Document made available under the Patent Cooperation Treaty (PCT)

International application number: PCT/US04/042980

International filing date: 23 December 2004 (23.12.2004)

Document type: Certified copy of priority document

Document details: Country/Office: US
Number: 60/532,205
Filing date: 23 December 2003 (23.12.2003)

Date of receipt at the International Bureau: 27 June 2005 (27.06.2005)

Remark: Priority document submitted or transmitted to the International Bureau in compliance with Rule 17.1(a) or (b)

1334635

THE UNITED STATES OF AMERICA

TO ALL TO WHOM THESE PRESENTS SHALL COME:

UNITED STATES DEPARTMENT OF COMMERCE

United States Patent and Trademark Office

June 16, 2005

THIS IS TO CERTIFY THAT ANNEXED HERETO IS A TRUE COPY FROM THE RECORDS OF THE UNITED STATES PATENT AND TRADEMARK OFFICE OF THOSE PAPERS OF THE BELOW IDENTIFIED PATENT APPLICATION THAT MET THE REQUIREMENTS TO BE GRANTED A FILING DATE.

APPLICATION NUMBER: ***60/532,205***
FILING DATE: ***December 23, 2003***
RELATED PCT APPLICATION NUMBER: ***PCT/US04/42980***

UNITED STATES PATENT AND TRADEMARK OFFICE
DEPARTMENT OF COMMERCE

Certified by

Jon W. Dudas

Under Secretary of Commerce
for Intellectual Property
and Director of the United States
Patent and Trademark Office

PTO/SB/16 (10-01)
Approved for use through10/31/2002. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

16569 U.S. PTO 122303

60/532205 U.S. PTO 122303

PROVISIONAL APPLICATION FOR PATENT COVER SHEET

This is a request for filing a PROVISIONAL APPLICATION FOR PATENT under 37 CFR 1.53(c).

Express Mail Label No.	EU404288861US

INVENTOR(S)

Given Name (first and middle [if any])	Family Name or Surname	Residence (City and either State or Foreign Country)
Neil	Berinstein	Toronto, Ontario, Canada
James	Tartaglia	Toronto, Ontario, Canada

☑ *Additional inventors are being named on the* 1 *separately numbered sheets attached hereto*

TITLE OF THE INVENTION (500 characters max)

MODIFIED KSA AND USES THEREOF

Direct all correspondence to: **CORRESPONDENCE ADDRESS**

☐ Customer Number ______ *Type Customer Number here* → *Place Customer Number Bar Code Label here*

OR

☑ Firm or Individual Name	Patrick J. Halloran				
Address	Aventis Pasteur, Inc.				
Address	Discovery Drive, Knerr Building				
City	Swiftwater	State	PA	ZIP	18370
Country	USA	Telephone	570-839-5446	Fax	570-895-2702

ENCLOSED APPLICATION PARTS ***(check all that apply)***

☑ Specification *Number of Pages* 39
☑ Drawing(s) *Number of Sheets* 22
☐ Application Data Sheet. See 37 CFR 1.76
☐ CD(s), Number
☐ Other (specify)

METHOD OF PAYMENT OF FILING FEES FOR THIS PROVISIONAL APPLICATION FOR PATENT

☐ Applicant claims small entity status. See 37 CFR 1.27.
☐ A check or money order is enclosed to cover the filing fees
☐ The Commissioner is hereby authorized to charge filing fees or credit any overpayment to Deposit Account Number:
☑ Payment by credit card. Form PTO-2038 is attached.

FILING FEE AMOUNT ($): $160.00

The invention was made by an agency of the United States Government or under a contract with an agency of the United States Government.

☑ No.
☐ Yes, the name of the U.S. Government agency and the Government contract number are: ______

Respectfully submitted,

SIGNATURE ______

TYPED or PRINTED NAME Patrick J. Halloran

TELEPHONE 570-839-5446

Date 12/23/2003

REGISTRATION NO. *(if appropriate)* 41,053

Docket Number: API-03-17-PR

USE ONLY FOR FILING A PROVISIONAL APPLICATION FOR PATENT

PROVISIONAL APPLICATION COVER SHEET
Additional Page

PTO/SB/16 (02-01)
Approved for use through 10/31/2002. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE

Docket Number	API-03-17-PR

INVENTOR(S)/APPLICANT(S)		
Given Name (first and middle [if any])	Family or Surname	Residence (City and either State or Foreign Country)
Mark	Parrington	Toronto, Ontario, Canada
Dennis	Panicalli	Boston, Massachusettes
Linda	Gritz	Boston, Massuchettes

Number 2 of 2

WARNING: Information on this form may become public. Credit card information should not be included on this form. Provide credit card information and authorization on PTO-2038.

U.S. Express Mail No. EU404288861US
Deposited December 23, 2003

MODIFIED KSA AND USES THEREOF

FIELD OF THE INVENTION

The present invention relates to a nucleic acid encoding a polypeptide and the use of the nucleic acid or polypeptide in preventing and / or treating cancer. In particular, the invention relates to improved vectors for the insertion and expression of foreign genes encoding tumor antigens for use in immunotherapeutic treatment of cancer.

BACKGROUND OF THE INVENTION

There has been tremendous increase in last few years in the development of cancer vaccines with Tumour-associated antigens (TAAs) due to the great advances in identification of molecules based on the expression profiling on primary tumours and normal cells with the help of several techniques such as high density microarray, SEREX, immunohistochemistry (IHC), RT-PCR, in-situ hybridization (ISH) and laser capture microscopy (Rosenberg, Immunity, 1999; Sgroi et al, 1999, Schena et al, 1995, Offringa et al, 2000). The TAAs are antigens expressed or over-expressed by tumour cells and could be specific to one or several tumours for example CEA antigen is expressed in colorectal, breast and lung cancers. Sgroi et al (1999) identified several genes differentially expressed in invasive and metastatic carcinoma cells with combined use of laser capture microdissection and cDNA microarrays. Several delivery systems like DNA or viruses could be used for therapeutic vaccination against human cancers (Bonnet et al, 2000) and can elicit immune responses and also break immune tolerance against TAAs. Tumour cells can be rendered more immunogenic by inserting transgenes encoding T cell co-stimulatory molecules such as B7.1 or cytokines IFNgamma, IL2, GM-CSF etc. Co-expression of a TAA and a cytokine or a co-stimulatory molecule can develop effective therapeutic vaccine (Hodge et al, 95, Bronte et al, 1995, Chamberlain et al, 1996).

There is a need in the art for reagents and methodologies useful in stimulating an immune response to prevent or treat cancers. The present inventions provides such reagents and methodologies which overcome many of the difficulties encountered by others in attempting to treat cancers such as cancer. In particular, the present invention provides an expression vector for expressing multiple tumor antigens and/or co-stimulatory components.

Such expression vectors are desired by those of skill in the art to improve anti-tumor immunity in cancer patients.

SUMMARY OF THE INVENTION

The present invention provides an immunogenic target for administration to a patient to prevent and / or treat cancer. In one embodiment, a single expression vector encoding the immunogenic targets CEA and p53 is provided (multiantigen expression vector). In another embodiment, a modified KSA sequence and vectors for expressing modified KSA are provided. Expression vectors encoding co-stimulatory components such as B7.1, LFA-3 and/or ICAM-1 in combination with CEA, p53 and/or KSA are also provided. In one embodiment, an ALVAC vector encoding CEA, p53, B7.1, LFA-3 and ICAM-1 is provided. In another embodiment, an ALVAC vector encoding modified KSA, B7.1, LFA-3 and ICAM-1 is provided. In yet another embodiment, an ALVAC vector encoding CEA, p53, modified KSA, B7.1, LFA-3 and ICAM-1 is provided. In certain embodiments, the expression vectors are administered to a patient as a nucleic acid contained within a plasmid or other delivery vector, such as a recombinant virus. The expression vector may also be administered in combination with an immune stimulator, such as a co-stimulatory molecule or adjuvant.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1. Donor plasmid useful in producing the ALVAC vector vcp2086.

Figure 2. Comparison of nucleotide sequence of CAP(6D) and CAP(6D)-1,2. Differences between the sequences are underlined.

Figure 3. A. Comparison of the amino acid sequences of wild-type KSA and modified KSA. **B.** DNA sequence encoding modified KSA

Figure 4. Construction of modified KSA plasmids.

Figure 5. A. Plasmid map of pT2255KSAV-1. **B.** DNA sequence of pT2255KSAV-1.

Figure 6. Plasmid maps of pALVAC.Tricom(C3)#33 and pT2255KSA(Val)LM.

DETAILED DESCRIPTION

The present invention provides reagents and methodologies useful for treating and / or preventing cancer. All references cited within this application are incorporated by reference.

In one embodiment, the present invention relates to the induction or enhancement of an immune response against one or more tumor antigens ("TA") to prevent and / or treat cancer. In certain embodiments, one or more TAs may be combined. In preferred embodiments, the immune response results from expression of a TA in a host cell following administration of a nucleic acid vector encoding the tumor antigen or the tumor antigen itself in the form of a peptide or polypeptide, for example.

As used herein, an "antigen" is a molecule (such as a polypeptide) or a portion thereof that produces an immune response in a host to whom the antigen has been administered. The immune response may include the production of antibodies that bind to at least one epitope of the antigen and / or the generation of a cellular immune response against cells expressing an epitope of the antigen. The response may be an enhancement of a current immune response by, for example, causing increased antibody production, production of antibodies with increased affinity for the antigen, or an increased cellular response (i.e., increased T cells). An antigen that produces an immune response may alternatively be referred to as being immunogenic or as an immunogen. In describing the present invention, a TA may be referred to as an "immunogenic target".

TA includes both tumor-associated antigens (TAAs) and tumor-specific antigens (TSAs), where a cancerous cell is the source of the antigen. A TAA is an antigen that is expressed on the surface of a tumor cell in higher amounts than is observed on normal cells or an antigen that is expressed on normal cells during fetal development. A TSA is an antigen that is unique to tumor cells and is not expressed on normal cells. TA further includes TAAs or TSAs, antigenic fragments thereof, and modified versions that retain their antigenicity.

TAs are typically classified into five categories according to their expression pattern, function, or genetic origin: cancer-testis (CT) antigens (i.e., MAGE, NY-ESO-1); melanocyte differentiation antigens (i.e., Melan A/MART-1, tyrosinase, gp100); mutational antigens (i.e., MUM-1, p53, CDK-4); overexpressed 'self' antigens (i.e., HER-2/neu, p53); and, viral antigens (i.e., HPV, EBV). For the purposes of practicing the present invention, a suitable TA is any TA that induces or enhances an anti-tumor immune response in a host to whom the TA has been administered. Suitable TAs include, for example, gp100 (Cox et al., *Science*, 264:716-719 (1994)), MART-1/Melan A (Kawakami et al., *J. Exp. Med.*, 180:347-352 (1994)), gp75 (TRP-1) (Wang et al., *J. Exp. Med.*, 186:1131-1140 (1996)), tyrosinase (Wolfel

et al., *Eur. J. Immunol.*, 24:759-764 (1994); WO 200175117; WO 200175016; WO 200175007), NY-ESO-1 (WO 98/14464; WO 99/18206), melanoma proteoglycan (Hellstrom et al., *J. Immunol.*, 130:1467-1472 (1983)), MAGE family antigens (i.e., MAGE-1, 2,3,4,6,12, 51; Van der Bruggen et al., *Science*, 254:1643-1647 (1991); U.S. Pat. Nos. 6,235,525; CN 1319611), BAGE family antigens (Boel et al., *Immunity*, 2:167-175 (1995)), GAGE family antigens (i.e., GAGE-1,2; Van den Eynde et al., *J. Exp. Med.*, 182:689-698 (1995); U.S. Pat. No. 6,013,765), RAGE family antigens (i.e., RAGE-1; Gaugler et at., *Immunogenetics*, 44:323-330 (1996); U.S. Pat. No. 5,939,526), N-acetylglucosaminyltransferase-V (Guilloux et at., *J. Exp. Med.*, 183:1173-1183 (1996)), p15 (Robbins et al., *J. Immunol.* 154:5944-5950 (1995)), ß-catenin (Robbins et al., *J. Exp. Med.*, 183:1185-1192 (1996)), MUM-1 (Coulie et al., *Proc. Natl. Acad. Sci.* USA, 92:7976-7980 (1995)), cyclin dependent kinase-4 (CDK4) (Wolfel et al., *Science*, 269:1281-1284 (1995)), p21-*ras* (Fossum et at., *Int. J. Cancer,* 56:40-45 (1994)), BCR-*abl* (Bocchia et al., *Blood*, 85:2680-2684 (1995)), p53 (Theobald et al., *Proc. Natl. Acad. Sci.* USA, 92:11993-11997 (1995)), p185 HER2/neu (erb-B1; Fisk et al., *J. Exp. Med.*, 181:2109-2117 (1995)), epidermal growth factor receptor (EGFR) (Harris et al., Breast Cancer Res. Treat, 29:1-2 (1994)), carcinoembryonic antigens (CEA) (Kwong et al., *J. Natl. Cancer Inst.*, 85:982-990 (1995) U.S. Pat. Nos. 5,756,103; 5,274,087; 5,571,710; 6,071,716; 5,698,530; 6,045,802; EP 263933; EP 346710; and, EP 784483); carcinoma-associated mutated mucins (i.e., MUC-1 gene products; Jerome et al., *J. Immunol.*, 151:1654-1662 (1993)); EBNA gene products of EBV (i.e., EBNA-1; Rickinson et al., *Cancer Surveys*, 13:53-80 (1992)); E7, E6 proteins of human papillomavirus (Ressing et al., *J. Immunol*, 154:5934-5943 (1995)); prostate specific antigen (PSA; Xue et al., *The Prostate*, 30:73-78 (1997)); prostate specific membrane antigen (PSMA; Israeli, et al., *Cancer Res.*, 54:1807-1811 (1994)); idiotypic epitopes or antigens, for example, immunoglobulin idiotypes or T cell receptor idiotypes (Chen et al., *J. Immunol.*, 153:4775-4787 (1994)); KSA (U.S. Patent No. 5,348,887), kinesin 2 (Dietz, et al. Biochem Biophys Res Commun 2000 Sep 7;275(3):731-8), HIP-55, TGFβ-1 anti-apoptotic factor (Toomey, et al. Br J Biomed Sci 2001;58(3):177-83), tumor protein D52 (Bryne J.A., et al., Genomics, 35:523-532 (1996)), H1FT, NY-BR-1 (WO 01/47959), NY-BR-62, NY-BR-75, NY-BR-85, NY-BR-87, NY-BR-96 (Scanlan, M. Serologic and Bioinformatic Approaches to the Identification of Human Tumor Antigens, in *Cancer Vaccines 2000*, Cancer Research Institute, New York, NY), including "wild-type" (i.e., normally encoded by

the genome, naturally-occurring), modified, and mutated versions as well as other fragments and derivatives thereof. Any of these TAs may be utilized alone or in combination with one another in a co-immunization protocol.

In certain cases, it may be beneficial to co-immunize patients with both TA and other antigens, such as angiogenesis-associated antigens ("AA"). An AA is an immunogenic molecule (i.e., peptide, polypeptide) associated with cells involved in the induction and / or continued development of blood vessels. For example, an AA may be expressed on an endothelial cell ("EC"), which is a primary structural component of blood vessels. Where the cancer is cancer, it is preferred that that the AA be found within or near blood vessels that supply a tumor. Immunization of a patient against an AA preferably results in an anti-AA immune response whereby angiogenic processes that occur near or within tumors are prevented and / or inhibited.

Exemplary AAs include, for example, vascular endothelial growth factor (i.e., VEGF; Bernardini, et al. *J. Urol.*, 2001, 166(4): 1275-9; Starnes, et al. *J. Thorac. Cardiovasc. Surg.*, 2001, 122(3): 518-23), the VEGF receptor (i.e., VEGF-R, flk-1/KDR; Starnes, et al. *J. Thorac. Cardiovasc. Surg.*, 2001, 122(3): 518-23), EPH receptors (i.e., EPHA2; Gerety, et al. 1999, *Cell*, 4: 403-414), epidermal growth factor receptor (i.e., EGFR; Ciardeillo, et al. Clin. Cancer Res., 2001, 7(10): 2958-70), basic fibroblast growth factor (i.e., bFGF; Davidson, et al. Clin. Exp. Metastasis 2000,18(6): 501-7; Poon, et al. Am J. Surg., 2001, 182(3):298-304), platelet-derived cell growth factor (i.e., PDGF-B), platelet-derived endothelial cell growth factor (PD-ECGF; Hong, et al. J. Mol. Med., 2001, 8(2):141-8), transforming growth factors (i.e., TGF-α; Hong, et al. J. Mol. Med., 2001, 8(2):141-8), endoglin (Balza, et al. *Int. J. Cancer*, 2001, 94: 579-585), Id proteins (Benezra, R. Trends Cardiovasc. Med., 2001, 11(6):237-41), proteases such as uPA, uPAR, and matrix metalloproteinases (MMP-2, MMP-9; Djonov, et al. J. Pathol., 2001, 195(2):147-55), nitric oxide synthase (Am. J. Ophthalmol., 2001, 132(4):551-6), aminopeptidase (Rouslhati, E. Nature Cancer, 2: 84-90, 2002), thrombospondins (i.e., TSP-1, TSP-2; Alvarez, et al. Gynecol. Oncol., 2001, 82(2):273-8; Seki, et al. Int. J. Oncol., 2001, 19(2):305-10), k-*ras* (Zhang, et al. Cancer Res., 2001, 61(16):6050-4), *Wnt* (Zhang, et al. Cancer Res., 2001, 61(16):6050-4), cyclin-dependent kinases (CDKs; Drug Resist. Updat. 2000, 3(2):83-88), microtubules (Timar, et al. 2001. *Path. Oncol. Res.*, 7(2): 85-94), heat shock proteins (i.e., HSP90 (Timar, *supra*)), heparin-binding factors (i.e., heparinase; Gohji, et al. Int. J. Cancer, 2001, 95(5):295-301), synthases

(i.e., ATP synthase, thymidilate synthase), collagen receptors, integrins (i.e., $\alpha\upsilon\beta3$, $\alpha\upsilon\beta5$, $\alpha1\beta1$, $\alpha2\beta1$, $\alpha5\beta1$), the surface proteolglycan NG2, AAC2-1, or AAC2-2, among others, including "wild-type" (i.e., normally encoded by the genome, naturally-occurring), modified, mutated versions as well as other fragments and derivatives thereof. Any of these targets may be suitable in practicing the present invention, either alone or in combination with one another or with other agents.

In certain embodiments, a nucleic acid molecule encoding an immunogenic target is utilized. The nucleic acid molecule may comprise or consist of a nucleotide sequence encoding one or more immunogenic targets, or fragments or derivatives thereof, such as that contained in a DNA insert in an ATCC Deposit. The term "nucleic acid sequence" or "nucleic acid molecule" refers to a DNA or RNA sequence. The term encompasses molecules formed from any of the known base analogs of DNA and RNA such as, but not limited to 4-acetylcytosine, 8-hydroxy-N6-methyladenosine, aziridinyl-cytosine, pseudoisocytosine, 5-(carboxyhydroxylmethyl) uracil, 5-fluorouracil, 5-bromouracil, 5-carboxymethylaminomethyl-2-thiouracil, 5-carboxy-methylaminomethyluracil, dihydrouracil, inosine, N6-iso-pentenyladenine, 1-methyladenine, 1-methylpseudouracil, 1-methylguanine, 1-methylinosine, 2,2-dimethyl-guanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-methyladenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyamino-methyl-2-thiouracil, beta-D-mannosylqueosine, 5' -methoxycarbonyl-methyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid, oxybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, N-uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid, pseudouracil, queosine, 2-thiocytosine, and 2,6-diaminopurine, among others.

An isolated nucleic acid molecule is one that: (1) is separated from at least about 50 percent of proteins, lipids, carbohydrates, or other materials with which it is naturally found when total nucleic acid is isolated from the source cells; (2) is not be linked to all or a portion of a polynucleotide to which the nucleic acid molecule is linked in nature; (3) is operably linked to a polynucleotide which it is not linked to in nature; and / or, (4) does not occur in nature as part of a larger polynucleotide sequence. Preferably, the isolated nucleic acid molecule of the present invention is substantially free from any other contaminating nucleic acid molecule(s) or other contaminants that are found in its natural environment that would

interfere with its use in polypeptide production or its therapeutic, diagnostic, prophylactic or research use. As used herein, the term "naturally occurring" or "native" or "naturally found" when used in connection with biological materials such as nucleic acid molecules, polypeptides, host cells, and the like, refers to materials which are found in nature and are not manipulated by man. Similarly, "non-naturally occurring" or "non-native" as used herein refers to a material that is not found in nature or that has been structurally modified or synthesized by man.

The identity of two or more nucleic acid or polypeptide molecules is determined by comparing the sequences. As known in the art, "identity" means the degree of sequence relatedness between nucleic acid molecules or polypeptides as determined by the match between the units making up the molecules (i.e., nucleotides or amino acid residues). Identity measures the percent of identical matches between the smaller of two or more sequences with gap alignments (if any) addressed by a particular mathematical model or computer program (i.e., an algorithm). Identity between nucleic acid sequences may also be determined by the ability of the related sequence to hybridize to the nucleic acid sequence or isolated nucleic acid molecule. In defining such sequences, the term "highly stringent conditions" and "moderately stringent conditions" refer to procedures that permit hybridization of nucleic acid strands whose sequences are complementary, and to exclude hybridization of significantly mismatched nucleic acids. Examples of "highly stringent conditions" for hybridization and washing are 0.015 M sodium chloride, 0.0015 M sodium citrate at 65-68°C or 0.015 M sodium chloride, 0.0015 M sodium citrate, and 50% formamide at 42°C. (see, for example, Sambrook, Fritsch & Maniatis, *Molecular Cloning: A Laboratory Manual* (2nd ed., Cold Spring Harbor Laboratory, 1989); Anderson *et al.*, *Nucleic Acid Hybridisation: A Practical Approach* Ch. 4 (IRL Press Limited)). The term "moderately stringent conditions" refers to conditions under which a DNA duplex with a greater degree of base pair mismatching than could occur under "highly stringent conditions" is able to form. Exemplary moderately stringent conditions are 0.015 M sodium chloride, 0.0015 M sodium citrate at 50-65°C or 0.015 M sodium chloride, 0.0015 M sodium citrate, and 20% formamide at 37-50°C. By way of example, moderately stringent conditions of 50°C in 0.015 M sodium ion will allow about a 21% mismatch. During hybridization, other agents may be included in the hybridization and washing buffers for the purpose of reducing non-specific and/or background hybridization. Examples are 0.1% bovine serum albumin, 0.1% polyvinyl-

pyrrolidone, 0.1% sodium pyrophosphate, 0.1% sodium dodecylsulfate, $NaDodSO_4$, (SDS), ficoll, Denhardt's solution, sonicated salmon sperm DNA (or another non-complementary DNA), and dextran sulfate, although other suitable agents can also be used. The concentration and types of these additives can be changed without substantially affecting the stringency of the hybridization conditions. Hybridization experiments are usually carried out at pH 6.8-7.4; however, at typical ionic strength conditions, the rate of hybridization is nearly independent of pH.

In preferred embodiments of the present invention, vectors are used to transfer a nucleic acid sequence encoding a polypeptide to a cell. A vector is any molecule used to transfer a nucleic acid sequence to a host cell. In certain cases, an expression vector is utilized. An expression vector is a nucleic acid molecule that is suitable for transformation of a host cell and contains nucleic acid sequences that direct and / or control the expression of the transferred nucleic acid sequences. Expression includes, but is not limited to, processes such as transcription, translation, and splicing, if introns are present. Expression vectors typically comprise one or more flanking sequences operably linked to a heterologous nucleic acid sequence encoding a polypeptide. Flanking sequences may be homologous (i.e., from the same species and / or strain as the host cell), heterologous (i.e., from a species other than the host cell species or strain), hybrid (i.e., a combination of flanking sequences from more than one source), or synthetic, for example.

A flanking sequence is preferably capable of effecting the replication, transcription and / or translation of the coding sequence and is operably linked to a coding sequence. As used herein, the term operably linked refers to a linkage of polynucleotide elements in a functional relationship. For instance, a promoter or enhancer is operably linked to a coding sequence if it affects the transcription of the coding sequence. However, a flanking sequence need not necessarily be contiguous with the coding sequence, so long as it functions correctly. Thus, for example, intervening untranslated yet transcribed sequences can be present between a promoter sequence and the coding sequence and the promoter sequence may still be considered operably linked to the coding sequence. Similarly, an enhancer sequence may be located upstream or downstream from the coding sequence and affect transcription of the sequence.

In certain embodiments, it is preferred that the flanking sequence is a trascriptional regulatory region that drives high-level gene expression in the target cell. The transcriptional

regulatory region may comprise, for example, a promoter, enhancer, silencer, repressor element, or combinations thereof. The transcriptional regulatory region may be either constitutive, tissue-specific, cell-type specific (i.e., the region is drives higher levels of transcription in a one type of tissue or cell as compared to another), or regulatable (i.e., responsive to interaction with a compound such as tetracycline). The source of a transcriptional regulatory region may be any prokaryotic or eukaryotic organism, any vertebrate or invertebrate organism, or any plant, provided that the flanking sequence functions in a cell by causing transcription of a nucleic acid within that cell. A wide variety of transcriptional regulatory regions may be utilized in practicing the present invention.

Suitable transcriptional regulatory regions include the CMV promoter (i.e., the CMV-immediate early promoter); promoters from eukaryotic genes (i.e., the estrogen-inducible chicken ovalbumin gene, the interferon genes, the gluco-corticoid-inducible tyrosine aminotransferase gene, and the thymidine kinase gene); and the major early and late adenovirus gene promoters; the SV40 early promoter region (Bernoist and Chambon, 1981, *Nature* 290:304-10); the promoter contained in the 3' long terminal repeat (LTR) of Rous sarcoma virus (RSV) (Yamamoto, *et al.*, 1980, *Cell* 22:787-97); the herpes simplex virus thymidine kinase (HSV-TK) promoter (Wagner *et al.*, 1981, *Proc. Natl. Acad. Sci. U.S.A.* 78:1444-45); the regulatory sequences of the metallothionine gene (Brinster *et al.*, 1982, *Nature* 296:39-42); prokaryotic expression vectors such as the beta-lactamase promoter (Villa-Kamaroff *et al.*, 1978, *Proc. Natl. Acad. Sci. U.S.A.*, 75:3727-31); or the tac promoter (DeBoer *et al.*, 1983, *Proc. Natl. Acad. Sci. U.S.A.*, 80:21-25). Tissue- and / or cell-type specific transcriptional control regions include, for example, the elastase I gene control region which is active in pancreatic acinar cells (Swift *et al.*, 1984, *Cell* 38:639-46; Ornitz *et al.*, 1986, *Cold Spring Harbor Symp. Quant. Biol.* 50:399-409 (1986); MacDonald, 1987, *Hepatology* 7:425-515); the insulin gene control region which is active in pancreatic beta cells (Hanahan, 1985, *Nature* 315:115-22); the immunoglobulin gene control region which is active in lymphoid cells (Grosschedl *et al.*, 1984, *Cell* 38:647-58; Adames *et al.*, 1985, *Nature* 318:533-38; Alexander *et al.*, 1987, *Mol. Cell. Biol.*, 7:1436-44); the mouse mammary tumor virus control region in testicular, breast, lymphoid and mast cells (Leder *et al.*, 1986, *Cell* 45:485-95); the albumin gene control region in liver (Pinkert *et al.*, 1987, *Genes and Devel.* 1:268-76); the alpha-feto-protein gene control region in liver (Krumlauf *et al.*, 1985, *Mol. Cell. Biol.*, 5:1639-48; Hammer *et al.*, 1987, *Science* 235:53-58); the alpha 1-

antitrypsin gene control region in liver (Kelsey *et al.*, 1987, *Genes and Devel.* 1:161-71); the beta-globin gene control region in myeloid cells (Mogram *et al.*, 1985, *Nature* 315:338-40; Kollias *et al.*, 1986, *Cell* 46:89-94); the myelin basic protein gene control region in oligodendrocyte cells in the brain (Readhead *et al.*, 1987, *Cell* 48:703-12); the myosin light chain-2 gene control region in skeletal muscle (Sani, 1985, *Nature* 314:283-86); the gonadotropic releasing hormone gene control region in the hypothalamus (Mason *et al.*, 1986, *Science* 234:1372-78), and the tyrosinase promoter in melanoma cells (Hart, I. Semin Oncol 1996 Feb;23(1):154-8; Siders, et al. Cancer Gene Ther 1998 Sep-Oct;5(5):281-91), among others. Other suitable promoters are known in the art.

As described above, enhancers may also be suitable flanking sequences. Enhancers are cis-acting elements of DNA, usually about 10-300 bp in length, that act on the promoter to increase transcription. Enhancers are typically orientation- and position-independent, having been identified both 5' and 3' to controlled coding sequences. Several enhancer sequences available from mammalian genes are known (i.e., globin, elastase, albumin, alpha-feto-protein and insulin). Similarly, the SV40 enhancer, the cytomegalovirus early promoter enhancer, the polyoma enhancer, and adenovirus enhancers are useful with eukaryotic promoter sequences. While an enhancer may be spliced into the vector at a position 5' or 3' to nucleic acid coding sequence, it is typically located at a site 5' from the promoter. Other suitable enhancers are known in the art, and would be applicable to the present invention.

While preparing reagents of the present invention, cells may need to be transfected or transformed. Transfection refers to the uptake of foreign or exogenous DNA by a cell, and a cell has been transfected when the exogenous DNA has been introduced inside the cell membrane. A number of transfection techniques are well known in the art (i.e., Graham *et al.*, 1973, *Virology* 52:456; Sambrook *et al.*, *Molecular Cloning, A Laboratory Manual* (Cold Spring Harbor Laboratories, 1989); Davis *et al.*, *Basic Methods in Molecular Biology* (Elsevier, 1986); and Chu *et al.*, 1981, *Gene* 13:197). Such techniques can be used to introduce one or more exogenous DNA moieties into suitable host cells.

In certain embodiments, it is preferred that transfection of a cell results in transformation of that cell. A cell is transformed when there is a change in a characteristic of the cell, being transformed when it has been modified to contain a new nucleic acid. Following transfection, the transfected nucleic acid may recombine with that of the cell by physically integrating into a chromosome of the cell, may be maintained transiently as an

episomal element without being replicated, or may replicate independently as a plasmid. A cell is stably transformed when the nucleic acid is replicated with the division of the cell.

The present invention further provides isolated immunogenic targets in polypeptide form. A polypeptide is considered isolated where it: (1) has been separated from at least about 50 percent of polynucleotides, lipids, carbohydrates, or other materials with which it is naturally found when isolated from the source cell; (2) is not linked (by covalent or noncovalent interaction) to all or a portion of a polypeptide to which the "isolated polypeptide" is linked in nature; (3) is operably linked (by covalent or noncovalent interaction) to a polypeptide with which it is not linked in nature; or, (4) does not occur in nature. Preferably, the isolated polypeptide is substantially free from any other contaminating polypeptides or other contaminants that are found in its natural environment that would interfere with its therapeutic, diagnostic, prophylactic or research use.

Immunogenic target polypeptides may be mature polypeptides, as defined herein, and may or may not have an amino terminal methionine residue, depending on the method by which they are prepared. Further contemplated are related polypeptides such as, for example, fragments, variants (i.e., allelic, splice), orthologs, homologues, and derivatives, for example, that possess at least one characteristic or activity (i.e., activity, antigenicity) of the immunogenic target. Also related are peptides, which refers to a series of contiguous amino acid residues having a sequence corresponding to at least a portion of the polypeptide from which its sequence is derived. In preferred embodiments, the peptide comprises about 5-10 amino acids, 10-15 amino acids, 15-20 amino acids, 20-30 amino acids, or 30-50 amino acids. In a more preferred embodiment, a peptide comprises 9-12 amino acids, suitable for presentation upon Class I MHC molecules, for example.

A fragment of a nucleic acid or polypeptide comprises a truncation of the sequence (i.e., nucleic acid or polypeptide) at the amino terminus (with or without a leader sequence) and / or the carboxy terminus. Fragments may also include variants (i.e., allelic, splice), orthologs, homologues, and other variants having one or more amino acid additions or substitutions or internal deletions as compared to the parental sequence. In preferred embodiments, truncations and/or deletions comprise about 10 amino acids, 20 amino acids, 30 amino acids, 40 amino acids, 50 amino acids, or more. The polypeptide fragments so produced will comprise about 10 amino acids, 25 amino acids, 30 amino acids, 40 amino acids, 50 amino acids, 60 amino acids, 70 amino acids, or more. Such polypeptide fragments

may optionally comprise an amino terminal methionine residue. It will be appreciated that such fragments can be used, for example, to generate antibodies or cellular immune responses to immunogenic target polypeptides.

A variant is a sequence having one or more sequence substitutions, deletions, and/or additions as compared to the subject sequence. Variants may be naturally occurring or artificially constructed. Such variants may be prepared from the corresponding nucleic acid molecules. In preferred embodiments, the variants have from 1 to 3, or from 1 to 5, or from 1 to 10, or from 1 to 15, or from 1 to 20, or from 1 to 25, or from 1 to 30, or from 1 to 40, or from 1 to 50, or more than 50 amino acid substitutions, insertions, additions and/or deletions.

An allelic variant is one of several possible naturally-occurring alternate forms of a gene occupying a given locus on a chromosome of an organism or a population of organisms. A splice variant is a polypeptide generated from one of several RNA transcript resulting from splicing of a primary transcript. An ortholog is a similar nucleic acid or polypeptide sequence from another species. For example, the mouse and human versions of an immunogenic target polypeptide may be considered orthologs of each other. A derivative of a sequence is one that is derived from a parental sequence those sequences having substitutions, additions, deletions, or chemically modified variants. Variants may also include fusion proteins, which refers to the fusion of one or more first sequences (such as a peptide) at the amino or carboxy terminus of at least one other sequence (such as a heterologous peptide).

"Similarity" is a concept related to identity, except that similarity refers to a measure of relatedness which includes both identical matches and conservative substitution matches. If two polypeptide sequences have, for example, 10/20 identical amino acids, and the remainder are all non-conservative substitutions, then the percent identity and similarity would both be 50%. If in the same example, there are five more positions where there are conservative substitutions, then the percent identity remains 50%, but the percent similarity would be 75% (15/20). Therefore, in cases where there are conservative substitutions, the percent similarity between two polypeptides will be higher than the percent identity between those two polypeptides.

Substitutions may be conservative, or non-conservative, or any combination thereof. Conservative amino acid modifications to the sequence of a polypeptide (and the corresponding modifications to the encoding nucleotides) may produce polypeptides having

functional and chemical characteristics similar to those of a parental polypeptide. For example, a "conservative amino acid substitution" may involve a substitution of a native amino acid residue with a non-native residue such that there is little or no effect on the size, polarity, charge, hydrophobicity, or hydrophilicity of the amino acid residue at that position and, in particlar, does not result in decreased immunogenicity. Suitable conservative amino acid substitutions are shown in **Table I**.

Table I

Original Residues	Exemplary Substitutions	Preferred Substitutions
Ala	Val, Leu, Ile	Val
Arg	Lys, Gln, Asn	Lys
Asn	Gln	Gln
Asp	Glu	Glu
Cys	Ser, Ala	Ser
Gln	Asn	Asn
Glu	Asp	Asp
Gly	Pro, Ala	Ala
His	Asn, Gln, Lys, Arg	Arg
Ile	Leu, Val, Met, Ala, Phe, Norleucine	Leu
Leu	Norleucine, Ile, Val, Met, Ala, Phe	Ile
Lys	Arg, 1,4 Diamino-butyric Acid, Gln, Asn	Arg
Met	Leu, Phe, Ile	Leu
Phe	Leu, Val, Ile, Ala, Tyr	Leu
Pro	Ala	Gly
Ser	Thr, Ala, Cys	Thr
Thr	Ser	Ser
Trp	Tyr, Phe	Tyr
Tyr	Trp, Phe, Thr, Ser	Phe
Val	Ile, Met, Leu, Phe, Ala, Norleucine	Leu

A skilled artisan will be able to determine suitable variants of polypeptide using well-known techniques. For identifying suitable areas of the molecule that may be changed without destroying biological activity (i.e., MHC binding, immunogenicity), one skilled in the art may target areas not believed to be important for that activity. For example, when similar polypeptides with similar activities from the same species or from other species are known, one skilled in the art may compare the amino acid sequence of a polypeptide to such similar polypeptides. By performing such analyses, one can identify residues and portions of the molecules that are conserved among similar polypeptides. It will be appreciated that changes in areas of the molecule that are not conserved relative to such similar polypeptides

would be less likely to adversely affect the biological activity and/or structure of a polypeptide. Similarly, the residues required for binding to MHC are known, and may be modified to improve binding. However, modifications resulting in decreased binding to MHC will not be appropriate in most situations. One skilled in the art would also know that, even in relatively conserved regions, one may substitute chemically similar amino acids for the naturally occurring residues while retaining activity. Therefore, even areas that may be important for biological activity or for structure may be subject to conservative amino acid substitutions without destroying the biological activity or without adversely affecting the polypeptide structure.

Other preferred polypeptide variants include glycosylation variants wherein the number and/or type of glycosylation sites have been altered compared to the subject amino acid sequence. In one embodiment, polypeptide variants comprise a greater or a lesser number of N-linked glycosylation sites than the subject amino acid sequence. An N-linked glycosylation site is characterized by the sequence Asn-X-Ser or Asn-X-Thr, wherein the amino acid residue designated as X may be any amino acid residue except proline. The substitution of amino acid residues to create this sequence provides a potential new site for the addition of an N-linked carbohydrate chain. Alternatively, substitutions that eliminate this sequence will remove an existing N-linked carbohydrate chain. Also provided is a rearrangement of N-linked carbohydrate chains wherein one or more N-linked glycosylation sites (typically those that are naturally occurring) are eliminated and one or more new N-linked sites are created. To affect O-linked glycosylation of a polypeptide, one would modify serine and / or threonine residues.

Additional preferred variants include cysteine variants, wherein one or more cysteine residues are deleted or substituted with another amino acid (*e.g.*, serine) as compared to the subject amino acid sequence set. Cysteine variants are useful when polypeptides must be refolded into a biologically active conformation such as after the isolation of insoluble inclusion bodies. Cysteine variants generally have fewer cysteine residues than the native protein, and typically have an even number to minimize interactions resulting from unpaired cysteines.

In other embodiments, the isolated polypeptides of the current invention include fusion polypeptide segments that assist in purification of the polypeptides. Fusions can be made either at the amino terminus or at the carboxy terminus of the subject polypeptide

variant thereof. Fusions may be direct with no linker or adapter molecule or may be through a linker or adapter molecule. A linker or adapter molecule may be one or more amino acid residues, typically from about 20 to about 50 amino acid residues. A linker or adapter molecule may also be designed with a cleavage site for a DNA restriction endonuclease or for a protease to allow for the separation of the fused moieties. It will be appreciated that once constructed, the fusion polypeptides can be derivatized according to the methods described herein. Suitable fusion segments include, among others, metal binding domains (e.g., a poly-histidine segment), immunoglobulin binding domains (i.e., Protein A, Protein G, T cell, B cell, Fc receptor, or complement protein antibody-binding domains), sugar binding domains (e.g., a maltose binding domain), and/or a "tag" domain (i.e., at least a portion of α-galactosidase, a strep tag peptide, a T7 tag peptide, a FLAG peptide, or other domains that can be purified using compounds that bind to the domain, such as monoclonal antibodies). This tag is typically fused to the polypeptide upon expression of the polypeptide, and can serve as a means for affinity purification of the sequence of interest polypeptide from the host cell. Affinity purification can be accomplished, for example, by column chromatography using antibodies against the tag as an affinity matrix. Optionally, the tag can subsequently be removed from the purified sequence of interest polypeptide by various means such as using certain peptidases for cleavage. As described below, fusions may also be made between a TA and a co-stimulatory components such as the chemokines CXC10 (IP-10), CCL7 (MCP-3), or CCL5 (RANTES), for example.

A fusion motif may enhance transport of an immunogenic target to an MHC processing compartment, such as the endoplasmic reticulum. These sequences, referred to as tranduction or transcytosis sequences, include sequences derived from HIV tat (see Kim et al. 1997 J. Immunol. 159:1666), *Drosophila* antennapedia (see Schutze-Redelmeier et al. 1996 J. Immunol. 157:650), or human period-1 protein (hPER1; in particular, SRRHHCRSKAKRSRHH).

In addition, the polypeptide or variant thereof may be fused to a homologous polypeptide to form a homodimer or to a heterologous polypeptide to form a heterodimer. Heterologous peptides and polypeptides include, but are not limited to: an epitope to allow for the detection and/or isolation of a fusion polypeptide; a transmembrane receptor protein or a portion thereof, such as an extracellular domain or a transmembrane and intracellular domain; a ligand or a portion thereof which binds to a transmembrane receptor protein; an

enzyme or portion thereof which is catalytically active; a polypeptide or peptide which promotes oligomerization, such as a leucine zipper domain; a polypeptide or peptide which increases stability, such as an immunoglobulin constant region; and a polypeptide which has a therapeutic activity different from the polypeptide or variant thereof.

In certain embodiments, it may be advantageous to combine a nucleic acid sequence encoding an immunogenic target, polypeptide, or derivative thereof with one or more co-stimulatory component(s) such as cell surface proteins, cytokines or chemokines in a composition of the present invention. The co-stimulatory component may be included in the composition as a polypeptide or as a nucleic acid encoding the polypeptide, for example. Suitable co-stimulatory molecules include, for instance, polypeptides that bind members of the CD28 family (i.e., CD28, ICOS; Hutloff, et al. *Nature* 1999, 397: 263–265; Peach, et al. *J Exp Med* 1994, 180: 2049–2058) such as the CD28 binding polypeptides B7.1 (CD80; Schwartz, 1992; Chen et al, 1992; Ellis, et al. *J. Immunol.*, 156(8): 2700-9) and B7.2 (CD86; Ellis, et al. *J. Immunol.*, 156(8): 2700-9); polypeptides which bind members of the integrin family (i.e., LFA-1 (CD11a / CD18); Sedwick, et al. *J Immunol* 1999, 162: 1367–1375; Wülfing, et al. *Science* 1998, 282: 2266–2269; Lub, et al. *Immunol Today* 1995, 16: 479–483) including members of the ICAM family (i.e., ICAM-1, -2 or -3); polypeptides which bind CD2 family members (i.e., CD2, signalling lymphocyte activation molecule (CDw150 or "SLAM"; Aversa, et al. *J Immunol* 1997, 158: 4036–4044)) such as CD58 (LFA-3; CD2 ligand; Davis, et al. *Immunol Today* 1996, 17: 177–187) or SLAM ligands (Sayos, et al. *Nature* 1998, 395: 462–469); polypeptides which bind heat stable antigen (HSA or CD24; Zhou, et al. *Eur J Immunol* 1997, 27: 2524–2528); polypeptides which bind to members of the TNF receptor (TNFR) family (i.e., 4-1BB (CD137; Vinay, et al. *Semin Immunol* 1998, 10: 481–489), OX40 (CD134; Weinberg, et al. *Semin Immunol* 1998, 10: 471–480; Higgins, et al. *J Immunol* 1999, 162: 486–493), and CD27 (Lens, et al. *Semin Immunol* 1998, 10: 491–499)) such as 4-1BBL (4-1BB ligand; Vinay, et al. *Semin Immunol* 1998, 10: 481–48; DeBenedette, et al. *J Immunol* 1997, 158: 551–559), TNFR associated factor-1 (TRAF-1; 4-1BB ligand; Saoulli, et al. *J Exp Med* 1998, 187: 1849–1862, Arch, et al. *Mol Cell Biol* 1998, 18: 558–565), TRAF-2 (4-1BB and OX40 ligand; Saoulli, et al. *J Exp Med* 1998, 187: 1849–1862; Oshima, et al. *Int Immunol* 1998, 10: 517–526, Kawamata, et al. *J Biol Chem* 1998, 273: 5808–5814), TRAF-3 (4-1BB and OX40 ligand; Arch, et al. *Mol Cell Biol* 1998,

18: 558–565; Jang, et al. *Biochem Biophys Res Commun* 1998, 242: 613–620; Kawamata S, et al. *J Biol Chem* 1998, 273: 5808–5814), OX40L (OX40 ligand; Gramaglia, et al. *J Immunol* 1998, 161: 6510–6517), TRAF-5 (OX40 ligand; Arch, et al. *Mol Cell Biol* 1998, 18: 558–565; Kawamata, et al. *J Biol Chem* 1998, 273: 5808–5814), and CD70 (CD27 ligand; Couderc, et al. *Cancer Gene Ther.*, 5(3): 163-75). CD154 (CD40 ligand or "CD40L"; Gurunathan, et al. *J. Immunol.*, 1998, 161: 4563-4571; Sine, et al. *Hum. Gene Ther.*, 2001, 12: 1091-1102) may also be suitable.

One or more cytokines may also be suitable co-stimulatory components or "adjuvants", either as polypeptides or being encoded by nucleic acids contained within the compositions of the present invention (Parmiani, et al. Immunol Lett 2000 Sep 15; 74(1): 41-4; Berzofsky, et al. Nature Immunol. 1: 209-219). Suitable cytokines include, for example, interleukin-2 (IL-2) (Rosenberg, et al. *Nature Med.* 4: 321-327 (1998)), IL-4, IL-7, IL-12 (reviewed by Pardoll, 1992; Harries, et al. J. Gene Med. 2000 Jul-Aug;2(4):243-9; Rao, et al. *J. Immunol.* 156: 3357-3365 (1996)), IL-15 (Xin, et al. *Vaccine*, 17:858-866, 1999), IL-16 (Cruikshank, et al. J. Leuk Biol. 67(6): 757-66, 2000), IL-18 (*J. Cancer Res. Clin. Oncol.* 2001. 127(12): 718-726), GM-CSF (CSF (Disis, et al. *Blood*, 88: 202-210 (1996)), or IFN.

As mentioned above, interferons may also be suitable cytokines for use in practicing the present invention. There are three main classes of interferon (alpha interferon (IFN-α), beta interferon (IFN-β) and gamma interferon (IFN-γ)) and at least 22 subtypes from among these. Many of these are available commercially. For instance, IFNs are commercially available as INFERGEN® (interferon alfacon-1; Intermune), Viraferon® (Schering-Plough), Roferon-A® (Roche) Wellferon® (Glaxo SmithKline), IFNα2b (Schering Canada, Pointe-Claire, Quebec), IFN beta-1b (Betaseron®; Berlex Laboratories), Avonex® (IFN beta-1a; Biogen); and Rebif® (IFN beta-1a ;Serono, Pfizer), Actimmune® (Interferon gamma-1b; Intermune). Preparations containing multiple IFN species in a single preparation are also available (i.e., IFN-alpha N3 or *Alferon N*). Variant and modified IFNs are also well-known (i.e., Maral, et al. Proc Am Soc Clin Oncol 22: page 174, 2003 (abstr 698); pegylated interferon alpha / Pegasys® (Roche); Peg Intron® (Schering Plough)). Other cytokines may also be suitable for practicing the present invention, as is known in the art. Other cytokines may also be suitable for practicing the present invention, as is known in the art.

Chemokines may also be utilized. For example, fusion proteins comprising CXCL10 (IP-10) and CCL7 (MCP-3) fused to a tumor self-antigen have been shown to induce anti-

tumor immunity (Biragyn, et al. *Nature Biotech.* 1999, 17: 253-258). The chemokines CCL3 (MIP-1α) and CCL5 (RANTES) (Boyer, et al. *Vaccine*, 1999, 17 (Supp. 2): S53-S64) may also be of use in practicing the present invention. Other suitable chemokines are known in the art.

It is also known in the art that suppressive or negative regulatory immune mechanisms may be blocked, resulting in enhanced immune responses. For instance, treatment with anti-CTLA-4 (Shrikant, et al. *Immunity*, 1996, 14: 145-155; Sutmuller, et al. *J. Exp. Med.*, 2001, 194: 823-832), anti-CD25 (Sutmuller, *supra*), anti-CD4 (Matsui, et al. *J. Immunol.*, 1999, 163: 184-193), the fusion protein IL13Ra2-Fc (Terabe, et al. *Nature Immunol.*, 2000, 1: 515-520), and combinations thereof (i.e., anti-CTLA-4 and anti-CD25, Sutmuller, *supra*) have been shown to upregulate anti-tumor immune responses and would be suitable in practicing the present invention.

Any of these components may be used alone or in combination with other agents. For instance, it has been shown that a combination of CD80, ICAM-1 and LFA-3 ("TRICOM") may potentiate anti-cancer immune responses (Hodge, et al. *Cancer Res.* 59: 5800-5807 (1999). Other effective combinations include, for example, IL-12 + GM-CSF (Ahlers, et al. *J. Immunol.*, 158: 3947-3958 (1997); Iwasaki, et al. *J. Immunol.* 158: 4591-4601 (1997)), IL-12 + GM-CSF + TNF-α (Ahlers, et al. *Int. Immunol.* 13: 897-908 (2001)), CD80 + IL-12 (Fruend, et al. *Int. J. Cancer*, 85: 508-517 (2000); Rao, et al. *supra*), and CD86 + GM-CSF + IL-12 (Iwasaki, supra). One of skill in the art would be aware of additional combinations useful in carrying out the present invention.In addition, the skilled artisan would be aware of additional reagents or methods that may be used to modulate such mechanisms. These reagents and methods, as well as others known by those of skill in the art, may be utilized in practicing the present invention.

Additional strategies for improving the efficiency of nucleic acid-based immunization may also be used including, for example, the use of self-replicating viral replicons (Caley, et al. 1999. *Vaccine*, 17: 3124-2135; Dubensky, et al. 2000. *Mol. Med.* 6: 723-732; Leitner, et al. 2000. *Cancer Res.* 60: 51-55), codon optimization (Liu, et al. 2000. *Mol. Ther.*, 1: 497-500; Dubensky, *supra*; Huang, et al. 2001. *J. Virol.* 75: 4947-4951), *in vivo* electroporation (Widera, et al. 2000. *J. Immunol.* 164: 4635-3640), incorporation of CpG stimulatory motifs (Gurunathan, et al. *Ann. Rev. Immunol.*, 2000, 18: 927-974; Leitner, *supra*), sequences for targeting of the endocytic or ubiquitin-processing pathways (Thomson, et al. 1998. *J. Virol.*

72: 2246-2252; Velders, et al. 2001. *J. Immunol.* 166: 5366-5373), prime-boost regimens (Gurunathan, *supra*; Sullivan, et al. 2000. *Nature*, 408: 605-609; Hanke, et al. 1998. *Vaccine*, 16: 439-445; Amara, et al. 2001. *Science*, 292: 69-74), and the use of mucosal delivery vectors such as *Salmonella* (Darji, et al. 1997. *Cell*, 91: 765-775; Woo, et al. 2001. *Vaccine*, 19: 2945-2954). Other methods are known in the art, some of which are described below.

Chemotherapeutic agents, radiation, anti-angiogenic compounds, or other agents may also be utilized in treating and / or preventing cancer using immunogenic targets (Sebti, et al. Oncogene 2000 Dec 27;19(56):6566-73). For example, in treating metastatic breast cancer, useful chemotherapeutic agents include cyclophosphamide, doxorubicin, paclitaxel, docetaxel, navelbine, capecitabine, and mitomycin C, among others. Combination chemotherapeutic regimens have also proven effective including cyclophosphamide + methotrexate + 5-fluorouracil; cyclophosphamide + doxorubicin + 5-fluorouracil; or, cyclophosphamide + doxorubicin, for example. Other compounds such as prednisone, a taxane, navelbine, mitomycin C, or vinblastine have been utlized for various reasons. A majority of breast cancer patients have estrogen-receptor positive (ER+) tumors and in these patients, endocrine therapy (i.e., tamoxifen) is preferred over chemotherapy. For such patients, tamoxifen or, as a second line therapy, progestins (medroxyprogesterone acetate or megestrol acetate) are preferred. Aromatase inhibitors (i.e., aminoglutethimide and analogs thereof such as letrozole) decrease the availability of estrogen needed to maintain tumor growth and may be used as second or third line endocrine therapy in certain patients.

Other cancers may require different chemotherapeutic regimens. For example, metastatic colorectal cancer is typically treated with Camptosar (irinotecan or CPT-11), 5-fluorouracil or leucovorin, alone or in combination with one another. Proteinase and integrin inhibitors such as as the MMP inhibitors marimastate (British Biotech), COL-3 (Collagenex), Neovastat (Aeterna), AG3340 (Agouron), BMS-275291 (Bristol Myers Squibb), CGS 27023A (Novartis) or the integrin inhibitors Vitaxin (Medimmune), or MED1522 (Merck KgaA) may also be suitable for use. As such, immunological targeting of immunogenic targets associated with colorectal cancer could be performed in combination with a treatment using those chemotherapeutic agents. Similarly, chemotherapeutic agents used to treat other types of cancers are well-known in the art and may be combined with the immunogenic targets described herein.

Many anti-angiogenic agents are known in the art and would be suitable for co-administration with the immunogenic target vaccines (see, for example, Timar, et al. 2001. *Pathology Oncol. Res.*, 7(2): 85-94). Such agents include, for example, physiological agents such as growth factors (i.e., ANG-2, NK1,2,4 (HGF), transforming growth factor beta (TGF-β)), cytokines (i.e., interferons such as IFN-α, -β, -γ, platelet factor 4 (PF-4), PR-39), proteases (i.e., cleaved AT-III, collagen XVIII fragment (Endostatin)), HmwKallikrein-d5 plasmin fragment (Angiostatin), prothrombin-F1-2, TSP-1), protease inhibitors (i.e., tissue inhibitor of metalloproteases such as TIMP-1, -2, or –3; maspin; plasminogen activator-inhibitors such as PAI-1; pigment epithelium derived factor (PEDF)), Tumstatin (available through ILEX, Inc.), antibody products (i.e., the collagen-binding antibodies HUIV26, HUI77, XL313; anti-VEGF; anti-integrin (i.e., Vitaxin, (Lxsys))), and glycosidases (i.e., heparinase-I, -III). "Chemical" or modified physiological agents known or believed to have anti-angiogenic potential include, for example, vinblastine, taxol, ketoconazole, thalidomide, dolestatin, combrestatin A, rapamycin (Guba, et al. 2002, *Nature Med.*, 8: 128-135), CEP-7055 (available from Cephalon, Inc.), flavone acetic acid, Bay 12-9566 (Bayer Corp.), AG3340 (Agouron, Inc.), CGS 27023A (Novartis), tetracylcine derivatives (i.e., COL-3 (Collagenix, Inc.)), Neovastat (Aeterna), BMS-275291 (Bristol-Myers Squibb), low dose 5-FU, low dose methotrexate (MTX), irsofladine, radicicol, cyclosporine, captopril, celecoxib, D45152-sulphated polysaccharide, cationic protein (Protamine), cationic peptide-VEGF, Suramin (polysulphonated napthyl urea), compounds that interfere with the function or production of VEGF (i.e., SU5416 or SU6668 (Sugen), PTK787/ZK22584 (Novartis)), Distamycin A, Angiozyme (ribozyme), isoflavinoids, staurosporine derivatives, genistein, EMD121974 (Merck KcgaA), tyrphostins, isoquinolones, retinoic acid, carboxyamidotriazole, TNP-470, octreotide, 2-methoxyestradiol, aminosterols (i.e., squalamine), glutathione analogues (i.e., N-acteyl-L-cysteine), combretastatin A-4 (Oxigene), Eph receptor blocking agents (*Nature*, 414:933-938, 2001), Rh-Angiostatin, Rh-Endostatin (WO 01/93897), cyclic-RGD peptide, accutin-disintegrin, benzodiazepenes, humanized anti-avb3 Ab, Rh-PAI-2, amiloride, p-amidobenzamidine, anti-uPA ab, anti-uPAR Ab, L-phanylalanin-N-methylamides (i.e., Batimistat, Marimastat), AG3340, and minocycline. Many other suitable agents are known in the art and would suffice in practicing the present invention.

The present invention may also be utilized in combination with "non-traditional" methods of treating cancer. For example, it has recently been demonstrated that administration of certain anaerobic bacteria may assist in slowing tumor growth. In one study, *Clostridium novyi* was modified to eliminate a toxin gene carried on a phage episome and administered to mice with colorectal tumors (Dang, et al. *P.N.A.S. USA*, 98(26): 15155-15160, 2001). In combination with chemotherapy, the treatment was shown to cause tumor necrosis in the animals. The reagents and methodologies described in this application may be combined with such treatment methodologies.

Nucleic acids encoding immunogenic targets may be administered to patients by any of several available techniques. Various viral vectors that have been successfully utilized for introducing a nucleic acid to a host include retrovirus, adenovirus, adeno-associated virus (AAV), herpes virus, and poxvirus, among others. It is understood in the art that many such viral vectors are available in the art. The vectors of the present invention may be constructed using standard recombinant techniques widely available to one skilled in the art. Such techniques may be found in common molecular biology references such as *Molecular Cloning: A Laboratory Manual* (Sambrook, et al., 1989, Cold Spring Harbor Laboratory Press), *Gene Expression Technology* (Methods in Enzymology, Vol. 185, edited by D. Goeddel, 1991. Academic Press, San Diego, CA), and *PCR Protocols: A Guide to Methods and Applications* (Innis, et al. 1990. Academic Press, San Diego, CA).

Preferred retroviral vectors are derivatives of lentivirus as well as derivatives of murine or avian retroviruses. Examples of suitable retroviral vectors include, for example, Moloney murine leukemia virus (MoMuLV), Harvey murine sarcoma virus (HaMuSV), murine mammary tumor virus (MuMTV), SIV, BIV, HIV and Rous Sarcoma Virus (RSV). A number of retroviral vectors can incorporate multiple exogenous nucleic acid sequences. As recombinant retroviruses are defective, they require assistance in order to produce infectious vector particles. This assistance can be provided by, for example, helper cell lines encoding retrovirus structural genes. Suitable helper cell lines include Ψ2, PA317 and PA12, among others. The vector virions produced using such cell lines may then be used to infect a tissue cell line, such as NIH 3T3 cells, to produce large quantities of chimeric retroviral virions. Retroviral vectors may be administered by traditional methods (i.e., injection) or by implantation of a "producer cell line" in proximity to the target cell population (Culver, K., *et al.*, 1994, *Hum. Gene Ther.*, 5 (3): 343-79; Culver, K., *et al.*, *Cold Spring Harb. Symp. Quant.*

Biol., 59: 685-90); Oldfield, E., 1993, *Hum. Gene Ther.,* 4 (1): 39-69). The producer cell line is engineered to produce a viral vector and releases viral particles in the vicinity of the target cell. A portion of the released viral particles contact the target cells and infect those cells, thus delivering a nucleic acid of the present invention to the target cell. Following infection of the target cell, expression of the nucleic acid of the vector occurs.

Adenoviral vectors have proven especially useful for gene transfer into eukaryotic cells (Rosenfeld, M., *et al.,* 1991, *Science,* 252 (5004): 431-4; Crystal, R., *et al.*, 1994, *Nat. Genet.,* 8 (1): 42-51), the study eukaryotic gene expression (Levrero, M., *et al.*, 1991, *Gene,* 101 (2): 195-202), vaccine development (Graham, F. and Prevec, L., 1992, *Biotechnology,* 20: 363-90), and in animal models (Stratford-Perricaudet, L., *et al.,* 1992, *Bone Marrow Transplant.,* 9 (Suppl. 1): 151-2 ; Rich, D., *et al.*, 1993, *Hum. Gene Ther.,* 4 (4): 461-76). Experimental routes for administrating recombinant Ad to different tissues *in vivo* have included intratracheal instillation (Rosenfeld, M., *et al.,* 1992, *Cell,* 68 (1): 143-55) injection into muscle (Quantin, B., *et al.,* 1992, *Proc. Natl. Acad. Sci. U.S.A.,* 89 (7): 2581-4), peripheral intravenous injection (Herz, J., and Gerard, R., 1993, *Proc. Natl. Acad. Sci. U.S.A.,* 90 (7): 2812-6) and stereotactic inoculation to brain (Le Gal La Salle, G., *et al.*, 1993, *Science,* 259 (5097): 988-90), among others.

Adeno-associated virus (AAV) demonstrates high-level infectivity, broad host range and specificity in integrating into the host cell genome (Hermonat, P., et al., 1984, *Proc. Natl. Acad. Sci. U.S.A.,* 81 (20): 6466-70). And Herpes Simplex Virus type-1 (HSV-1) is yet another attractive vector system, especially for use in the nervous system because of its neurotropic property (Geller, A., *et al.,* 1991, *Trends Neurosci.,* 14 (10): 428-32; Glorioso, *et al.*, 1995, *Mol. Biotechnol.,* 4 (1): 87-99; Glorioso, *et al.*, 1995, *Annu. Rev. Microbiol.,* 49: 675-710).

Poxvirus is another useful expression vector (Smith, et al. 1983, *Gene,* 25 (1): 21-8; Moss, et al, 1992, *Biotechnology,* 20: 345-62; Moss, et al, 1992, *Curr. Top. Microbiol. Immunol.*, 158: 25-38; Moss, et al. 1991. *Science*, 252: 1662-1667). Poxviruses shown to be useful include vaccinia, NYVAC, avipox, fowlpox, canarypox, ALVAC, and ALVAC(2), among others.

Vaccinia virus is the prototypic virus of the pox virus family and, like other members of the pox virus group, is distinguished by its large size and complexity. The DNA of vaccinia virus is similarly large and complex. Several types of vaccinia are suitable for use in

practicing the present invention. One such vaccinia-related virus is the Modified Vaccinia Virus Ankara (MVA), as described in, for example, U.S. Pat. Nos. 5,185,146 and 6,440,422.

Another suitable vaccinia-related virus is NYVAC. NYVAC was derived from the Copenhagen vaccine strain of vaccinia virus by deleting six nonessential regions of the genome encoding known or potential virulence factors (see, for example, U.S. Pat. Nos. 5,364,773 and 5,494,807). The deletion loci were also engineered as recipient loci for the insertion of foreign genes. The deleted regions are: thymidine kinase gene (TK; J2R); hemorrhagic region (u; B13R+B14R); A type inclusion body region (ATI; A26L); hemagglutinin gene (HA; A56R); host range gene region (C7L-K1L); and, large subunit, ribonucleotide reductase (I4L). NYVAC is a genetically engineered vaccinia virus strain that was generated by the specific deletion of eighteen open reading frames encoding gene products associated with virulence and host range. NYVAC has been show to be useful for expressing TAs (see, for example, U.S. Pat. No. 6,265,189). NYVAC (vP866), vP994, vCP205, vCP1433, placZH6H4Lreverse, pMPC6H6K3E3 and pC3H6FHVB were also deposited with the ATCC under the terms of the Budapest Treaty, accession numbers VR-2559, VR-2558, VR-2557, VR-2556, ATCC-97913, ATCC-97912, and ATCC-97914, respectively.

ALVAC-based recombinant viruses (i.e., ALVAC-1 and ALVAC-2) are also suitable for use in practicing the present invention (see, for example, U.S. Pat. No. 5,756,103). ALVAC(2) is identical to ALVAC(1) except that ALVAC(2) genome comprises the vaccinia E3L and K3L genes under the control of vaccinia promoters (U.S. Pat. No. 6,130,066; Beattie et al., 1995a, 1995b, 1991; Chang et al., 1992; Davies et al., 1993). Both ALVAC(1) and ALVAC(2) have been demonstrated to be useful in expressing foreign DNA sequences, such as TAs (Tartaglia et al., 1993 a,b; U.S. Pat. No. 5,833,975). ALVAC was deposited under the terms of the Budapest Treaty with the American Type Culture Collection (ATCC), 10801 University Boulevard, Manassas, Va. 20110-2209, USA, ATCC accession number VR-2547.

Another useful poxvirus vector is TROVAC. TROVAC refers to an attenuated fowlpox that was a plaque-cloned isolate derived from the FP-1 vaccine strain of fowlpoxvirus which is licensed for vaccination of 1 day old chicks. TROVAC was likewise deposited under the terms of the Budapest Treaty with the ATCC, accession number 2553.

"Non-viral" plasmid vectors may also be suitable in practicing the present invention. Preferred plasmid vectors are compatible with bacterial, insect, and / or mammalian host

cells. Such vectors include, for example, PCR-II, pCR3, and pcDNA3.1 (Invitrogen, San Diego, CA), pBSII (Stratagene, La Jolla, CA), pET15 (Novagen, Madison, WI), pGEX (Pharmacia Biotech, Piscataway, NJ), pEGFP-N2 (Clontech, Palo Alto, CA), pETL (BlueBacII, Invitrogen), pDSR-alpha (PCT pub. No. WO 90/14363) and pFastBacDual (Gibco-BRL, Grand Island, NY) as well as Bluescript® plasmid derivatives (a high copy number COLE1-based phagemid, Stratagene Cloning Systems, La Jolla, CA), PCR cloning plasmids designed for cloning Taq-amplified PCR products (*e.g.*, TOPO™ TA cloning® kit, PCR2.1® plasmid derivatives, Invitrogen, Carlsbad, CA). Bacterial vectors may also be used with the current invention. These vectors include, for example, *Shigella, Salmonella, Vibrio cholerae, Lactobacillus, Bacille calmette guérin (BCG); and Streptococcus* (see for example, WO 88/6626; WO 90/0594; WO 91/13157; WO 92/1796; and WO 92/21376). Many other non-viral plasmid expression vectors and systems are known in the art and could be used with the current invention.

Suitable nucleic acid delivery techniques include DNA-ligand complexes, adenovirus-ligand-DNA complexes, direct injection of DNA, $CaPO_4$ precipitation, gene gun techniques, electroporation, and colloidal dispersion systems, among others. Colloidal dispersion systems include macromolecule complexes, nanocapsules, microspheres, beads, and lipid-based systems including oil-in-water emulsions, micelles, mixed micelles, and liposomes. The preferred colloidal system of this invention is a liposome, which are artificial membrane vesicles useful as delivery vehicles *in vitro* and *in vivo*. RNA, DNA and intact virions can be encapsulated within the aqueous interior and be delivered to cells in a biologically active form (Fraley, R., *et al.,* 1981, *Trends Biochem. Sci.*, 6: 77). The composition of the liposome is usually a combination of phospholipids, particularly high-phase-transition-temperature phospholipids, usually in combination with steroids, especially cholesterol. Other phospholipids or other lipids may also be used. The physical characteristics of liposomes depend on pH, ionic strength, and the presence of divalent cations. Examples of lipids useful in liposome production include phosphatidyl compounds, such as phosphatidylglycerol, phosphatidylcholine, phosphatidylserine, phosphatidylethanolamine, sphingolipids, cerebrosides, and gangliosides. Particularly useful are diacylphosphatidylglycerols, where the lipid moiety contains from 14-18 carbon atoms, particularly from 16-18 carbon atoms, and is saturated. Illustrative phospholipids include egg phosphatidylcholine, dipalmitoylphosphatidylcholine and distearoylphosphatidylcholine.

An immunogenic target may also be administered in combination with one or more adjuvants to boost the immune response. Exemplary adjuvants are shown in Table II below:

Table II

Types of Immunologic Adjuvants

	Type of Adjuvant	General Examples	Specific Examples/References
1	Gel-type	Aluminum hydroxide/phosphate ("alum adjuvants")	(Aggerbeck and Heron, 1995)
		Calcium phosphate	(Relyveld, 1986)
2	Microbial	Muramyl dipeptide (MDP)	(Chedid et al., 1986)
		Bacterial exotoxins	Cholera toxin (CT), *E.coli* labile toxin (LT)(Freytag and Clements, 1999)
		Endotoxin-based adjuvants	Monophosphoryl lipid A (MPL) (Ulrich and Myers, 1995)
		Other bacterial	CpG oligonucleotides (Corral and Petray, 2000), BCG sequences (Krieg, et al. *Nature*, 374:576), tetanus toxoid (Rice, et al. *J. Immunol.*, 2001, 167: 1558-1565)
3	Particulate	Biodegradable polymer microspheres	(Gupta et al., 1998)
		Immunostimulatory complexes (ISCOMs)	(Morein and Bengtsson, 1999)
		Liposomes	(Wassef et al., 1994)
4	Oil-emulsion and surfactant-based adjuvants	Freund's incomplete adjuvant	(Jensen et al., 1998)
		Microfluidized emulsions	MF59 (Ott et al., 1995)
			SAF (Allison and Byars, 1992) (Allison, 1999)
		Saponins	QS-21 (Kensil, 1996)
5	Synthetic	Muramyl peptide derivatives	Murabutide (Lederer, 1986) Threony-MDP (Allison, 1997)
		Nonionic block copolymers	L121 (Allison, 1999)
		Polyphosphazene (PCPP)	(Payne et al., 1995)
		Synthetic polynucleotides	Poly A:U, Poly I:C (Johnson, 1994)

The immunogenic targets of the present invention may also be used to generate antibodies for use in screening assays or for immunotherapy. Other uses would be apparent to one of skill in the art. The term "antibody" includes antibody fragments, as are known in the art, including Fab, Fab_2, single chain antibodies (Fv for example), humanized antibodies, chimeric antibodies, human antibodies, produced by several methods as are known in the art. Methods of preparing and utilizing various types of antibodies are well-known to those of

skill in the art and would be suitable in practicing the present invention (see, for example, Harlow, et al. *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory, 1988; Harlow, et al. *Using Antibodies: A Laboratory Manual, Portable Protocol No. 1*, 1998; Kohler and Milstein, Nature, 256:495 (1975)); Jones et al. Nature, 321:522-525 (1986); Riechmann et al. Nature, 332:323-329 (1988); Presta (Curr. Op. Struct. Biol., 2:593-596 (1992); Verhoeyen et al. (Science, 239:1534-1536 (1988); Hoogenboom et al., J. Mol. Biol., 227:381 (1991); Marks et al., J. Mol. Biol., 222:581 (1991); Cole et al., Monoclonal Antibodies and Cancer Therapy, Alan R. Liss, p. 77 (1985); Boerner et al., J. Immunol., 147(1):86-95 (1991); Marks et al., Bio/Technology 10, 779-783 (1992); Lonberg et al., Nature 368 856-859 (1994); Morrison, Nature 368 812-13 (1994); Fishwild et al., Nature Biotechnology 14, 845-51 (1996); Neuberger, Nature Biotechnology 14, 826 (1996); Lonberg and Huszar, Intern. Rev. Immunol. 13 65-93 (1995); as well as U.S. Pat. Nos. 4,816,567; 5,545,807; 5,545,806; 5,569,825; 5,625,126; 5,633,425; and, 5,661,016). The antibodies or derivatives therefrom may also be conjugated to therapeutic moieties such as cytotoxic drugs or toxins, or active fragments thereof such as diptheria A chain, exotoxin A chain, ricin A chain, abrin A chain, curcin, crotin, phenomycin, enomycin, among others. Cytotoxic agents may also include radiochemicals. Antibodies and their derivatives may be incorporated into compositions of the invention for use *in vitro* or *in vivo*.

Nucleic acids, proteins, or derivatives thereof representing an immunogenic target may be used in assays to determine the presence of a disease state in a patient, to predict prognosis, or to determine the effectiveness of a chemotherapeutic or other treatment regimen. Expression profiles, performed as is known in the art, may be used to determine the relative level of expression of the immunogenic target. The level of expression may then be correlated with base levels to determine whether a particular disease is present within the patient, the patient's prognosis, or whether a particular treatment regimen is effective. For example, if the patient is being treated with a particular chemotherapeutic regimen, an decreased level of expression of an immunogenic target in the patient's tissues (i.e., in peripheral blood) may indicate the regimen is decreasing the cancer load in that host. Similarly, if the level of expresssion is increasing, another therapeutic modality may need to be utilized. In one embodiment, nucleic acid probes corresponding to a nucleic acid encoding an immunogenic target may be attached to a biochip, as is known in the art, for the detection and quantification of expression in the host.

It is also possible to use nucleic acids, proteins, derivatives therefrom, or antibodies thereto as reagents in drug screening assays. The reagents may be used to ascertain the effect of a drug candidate on the expression of the immunogenic target in a cell line, or a cell or tissue of a patient. The expression profiling technique may be combined with high throughput screening techniques to allow rapid identification of useful compounds and monitor the effectiveness of treatment with a drug candidate (see, for example, Zlokarnik, et al., Science 279, 84-8 (1998)). Drug candidates may be chemical compounds, nucleic acids, proteins, antibodies, or derivatives therefrom, whether naturally occurring or synthetically derived. Drug candidates thus identified may be utilized, among other uses, as pharmaceutical compositions for administration to patients or for use in further screening assays.

Administration of a composition of the present invention to a host may be accomplished using any of a variety of techniques known to those of skill in the art. The composition(s) may be processed in accordance with conventional methods of pharmacy to produce medicinal agents for administration to patients, including humans and other mammals (i.e., a "pharmaceutical composition"). The pharmaceutical composition is preferably made in the form of a dosage unit containing a given amount of DNA, viral vector particles, polypeptide or peptide, for example. A suitable daily dose for a human or other mammal may vary widely depending on the condition of the patient and other factors, but, once again, can be determined using routine methods.

The pharmaceutical composition may be administered orally, parentally, by inhalation spray, rectally, intranodally, or topically in dosage unit formulations containing conventional pharmaceutically acceptable carriers, adjuvants, and vehicles. The term "pharmaceutically acceptable carrier" or "physiologically acceptable carrier" as used herein refers to one or more formulation materials suitable for accomplishing or enhancing the delivery of a nucleic acid, polypeptide, or peptide as a pharmaceutical composition. A "pharmaceutical composition" is a composition comprising a therapeutically effective amount of a nucleic acid or polypeptide. The terms "effective amount" and "therapeutically effective amount" each refer to the amount of a nucleic acid or polypeptide used to induce or enhance an effective immune response. It is preferred that compositions of the present invention provide for the induction or enhancement of an anti-tumor immune response in a host which protects

the host from the development of a tumor and / or allows the host to eliminate an existing tumor from the body.

For oral administration, the pharmaceutical composition may be of any of several forms including, for example, a capsule, a tablet, a suspension, or liquid, among others. Liquids may be administered by injection as a composition with suitable carriers including saline, dextrose, or water. The term parenteral as used herein includes subcutaneous, intravenous, intramuscular, intrasternal, infusion, or intraperitoneal administration. Suppositories for rectal administration of the drug can be prepared by mixing the drug with a suitable non-irritating excipient such as cocoa butter and polyethylene glycols that are solid at ordinary temperatures but liquid at the rectal temperature.

The dosage regimen for immunizing a host or otherwise treating a disorder or a disease with a composition of this invention is based on a variety of factors, including the type of disease, the age, weight, sex, medical condition of the patient, the severity of the condition, the route of administration, and the particular compound employed. For example, a poxviral vector may be administered as a composition comprising 1 x 10^6 infectious particles per dose. Thus, the dosage regimen may vary widely, but can be determined routinely using standard methods.

A prime-boost regimen may also be utilized (WO 01/30382 A1) in which the targeted immunogen is initially administered in a priming step in one form followed by a boosting step in which the targeted immunogen is administered in another form. The form of the targeted immunogen in the priming and boosting steps are different. For instance, if the priming step utilized a nucleic acid, the boost may be administered as a peptide. Simmilarly, where a priming step utilized one type of recombinant virus (i.e., ALVAC), the boost step may utilize another type of virus (i.e., NYVAC). This prime-boost method of administration has been shown to induce strong immunological responses.

While the compositions of the invention can be administered as the sole active pharmaceutical agent, they can also be used in combination with one or more other compositions or agents (i.e., other immunogenic targets, co-stimulatory molecules, adjuvants). When administered as a combination, the individual components can be formulated as separate compositions administered at the same time or different times, or the components can be combined as a single composition.

Injectable preparations, such as sterile injectable aqueous or oleaginous suspensions, may be formulated according to known methods using suitable dispersing or wetting agents and suspending agents. The injectable preparation may also be a sterile injectable solution or suspension in a non-toxic parenterally acceptable diluent or solvent. Suitable vehicles and solvents that may be employed are water, Ringer's solution, and isotonic sodium chloride solution, among others. For instance, a viral vector such as a poxvirus may be prepared in 0.4% NaCl. In addition, sterile, fixed oils are conventionally employed as a solvent or suspending medium. For this purpose, any bland fixed oil may be employed, including synthetic mono- or diglycerides. In addition, fatty acids such as oleic acid find use in the preparation of injectables.

For topical administration, a suitable topical dose of a composition may be administered one to four, and preferably two or three times daily. The dose may also be administered with intervening days during which no does is applied. Suitable compositions may comprise from 0.001% to 10% w/w, for example, from 1% to 2% by weight of the formulation, although it may comprise as much as 10% w/w, but preferably not more than 5% w/w, and more preferably from 0.1% to 1% of the formulation. Formulations suitable for topical administration include liquid or semi-liquid preparations suitable for penetration through the skin (*e.g.*, liniments, lotions, ointments, creams, or pastes) and drops suitable for administration to the eye, ear, or nose.

The pharmaceutical compositions may also be prepared in a solid form (including granules, powders or suppositories). The pharmaceutical compositions may be subjected to conventional pharmaceutical operations such as sterilization and/or may contain conventional adjuvants, such as preservatives, stabilizers, wetting agents, emulsifiers, buffers etc. Solid dosage forms for oral administration may include capsules, tablets, pills, powders, and granules. In such solid dosage forms, the active compound may be admixed with at least one inert diluent such as sucrose, lactose, or starch. Such dosage forms may also comprise, as in normal practice, additional substances other than inert diluents, *e.g.*, lubricating agents such as magnesium stearate. In the case of capsules, tablets, and pills, the dosage forms may also comprise buffering agents. Tablets and pills can additionally be prepared with enteric coatings. Liquid dosage forms for oral administration may include pharmaceutically acceptable emulsions, solutions, suspensions, syrups, and elixirs containing inert diluents

commonly used in the art, such as water. Such compositions may also comprise adjuvants, such as wetting sweetening, flavoring, and perfuming agents.

Pharmaceutical compositions comprising a nucleic acid or polypeptide of the present invention may take any of several forms and may be administered by any of several routes. In preferred embodiments, the compositions are administered via a parenteral route (intradermal, intramuscular or subcutaneous) to induce an immune response in the host. Alternatively, the composition may be administered directly into a lymph node (intranodal) or tumor mass (i.e., intratumoral administration). For example, the dose could be administered subcutaneously at days 0, 7, and 14. Suitable methods for immunization using compositions comprising TAs are known in the art, as shown for p53 (Hollstein et al., 1991), p21-ras (Almoguera et al., 1988), HER-2 (Fendly et al., 1990), the melanoma-associated antigens (MAGE-1; MAGE-2) (van der Bruggen et al., 1991), p97 (Hu et al., 1988), and carcinoembryonic antigen (CEA) (Kantor et al., 1993; Fishbein et al., 1992; Kaufman et al., 1991), among others.

Preferred embodiments of administratable compositions include, for example, nucleic acids or polypeptides in liquid preparations such as suspensions, syrups, or elixirs. Preferred injectable preparations include, for example, nucleic acids or polypeptides suitable for parental, subcutaneous, intradermal, intramuscular or intravenous administration such as sterile suspensions or emulsions. For example, a recombinant poxvirus may be in admixture with a suitable carrier, diluent, or excipient such as sterile water, physiological saline, glucose or the like. The composition may also be provided in lyophilized form for reconstituting, for instance, in isotonic aqueous, saline buffer.. In addition, the compositions can be co-administered or sequentially administered with other antineoplastic, anti-tumor or anti-cancer agents and/or with agents which reduce or alleviate ill effects of antineoplastic, anti-tumor or anti-cancer agents.

A kit comprising a composition of the present invention is also provided. The kit can include a separate container containing a suitable carrier, diluent or excipient. The kit can also include an additional anti-cancer, anti-tumor or antineoplastic agent and/or an agent that reduces or alleviates ill effects of antineoplastic, anti-tumor or anti-cancer agents for co- or sequential-administration. Additionally, the kit can include instructions for mixing or combining ingredients and/or administration.

A better understanding of the present invention and of its many advantages will be had from the following examples, given by way of illustration.

EXAMPLES

Example 1

Vectors

A. Construction of the Multi-Antigen Construct vcp2086

An expression vector was constructed in the ALVAC(2) vector using standard techniques. DNA sequences encoding LFA-3 (Wallner, et al. (1987) J. Exp. Med. 166:923-932), ICAM-1 (Staunton, et al. (1988) Cell 52:925-933) and B7.1 (Chen, et al. (1992) Cell 71:1093-1102) were inserted into the C3 locus of ALVAC. LFA-3, ICAM-1 and B7.1 form an expression cassette known as TRICOM. DNA sequences encoding CEA-CAP1(6D) and p53 were inserted into the ALVAC donor plasmid pNC5LSPCEAp53 as shown in **Figure 1**. This donor plasmid was then used with the ALVAC-TRICOM vector to generate vcp2086 (ALVAC-CEA-p53-TRICOM).

B. Construction of the Multi-Antigen Construct Containing CEA-CAP1-6D-1,2

An expression vector is constructed in the ALVAC(2) vector using standard techniques. DNA sequences encoding LFA-3 (Wallner, et al. (1987) J. Exp. Med. 166:923-932), ICAM-1 (Staunton, et al. (1988) Cell 52:925-933) and B7.1 (Chen, et al. (1992) Cell 71:1093-1102) are inserted into the C3 locus of ALVAC. LFA-3, ICAM-1 and B7.1 form an expression cassette known as TRICOM. DNA sequences encoding CEA-CAP1(6D)-1,2 (**Fig. 2**) and p53 are inserted into the ALVAC donor plasmid essentially as shown in **Figure 1**. In this vector, CEA-CAP1-6D is removed and CEA-CAP1-6D-1,2 (**Fig. 2**) is inserted using standard techniques. This donor plasmid was then used with the ALVAC-TRICOM vector to generate vcp2086 (ALVAC-CEA-p53-TRICOM).

EXAMPLE 2

Immunogenicity of Multiantigen Vectors

This series of experiments was designed to confirm the immunogenicity of the multiantigen expression vectors. As an example, vcp2086 was administered to the double transgenic mouse strain "CEA/A2K^{b}dbTg". These mice express both the chimeric HLA.A2kb Class I molecule as well as the human CEA gene as a "self" antigen. The potential to generate strong immunogenicity in this model depends upon the ability of the expression vectors to break tolerance and generate a T cell response to the self antigen CEA.

Detection of anti-p53 responses is evaluated in the context of p53 being a foreign antigen, and therefore the issue of tolerance may not apply to p53 in this model.

A. Study MAD68

This experiment was designed as a dose titer of the multiantigen constructs. As a vector control, animals were immunized with the ALVAC(2) parental vector over an identical dose range. Analysis of immunogenicity is based on an ELIPSOT assay to detect IFN-γ production by peptide-specific T cells present in cultures from individual CEAxHLA.A2Kb Tg mice immunized with the indicated recombinant viruses. Groups of three individual mice were tested for each recombinant at a particular dose. Replicate cultures for all data points were tested against a control peptide to determine background response levels of the ELISPOT assay. The average of the three individual mice in each group was determined for comparison between groups. As a positive control, each individual culture group was tested using the mitogens PMA/ionomycin to induce IFN-γ from total spleen cells.

Individual spleen cells from the different groups (vcp2086 or ALVAC(2) parental vector at $1x10^8$; $2x10^7$; $2x10^6$; $2x10^5$ pfu/mouse) were harvested and re-stimulated *in vitro* with CEA or p53 peptides (**Table III**).

TABLE III

CEA and p53 Peptides

Peptide	Internal ID	Amino Acid Sequence
CEA-24	3205	LLTFWNPPT
CEA-233	1815	VLYGPDAPTI
CEA-691	571	IMIGVLVGV
CEA-78	3209	QIIGYVIGT
P53-139-147	3211	KTCPVQLWV
P53-149-157	3213	STPPPGTRV
P53-101-111	3215	KTYQGSYGFRL
P53-216	3217	VVVPYEPPEV

Duplicate bulk cultures were stimulated *in vitro* in a second round with peptide pulsed activated B cells. At the 2 x 10^5 pfu/mouse, responses above parental control vector reactivity was observed following separate stimulation with peptides CEA-78, CEA-233, CEA-591, p53-101, and p53-216. The strongest responses were detected using CEA-233 or p53-216.

Intracellular cytokine staining (ICS) was performed following stimulation with the most reactive epitopes (CEA-233 and p53-216). The percent positive CD8+ lymphocytes was increased relative to control at the 2 x 10^5 pfu/mouse dose level for both CEA-233 and p53-216.

CTL activity was also measured following immunization of CEA/HLA.A2kb mice with vcp2086 (ALVAC-CEA-p53-TRICOM) or the parental ALVAC(2) vector. The following immunization protocol was utilized. On day 0, animals were administered $2x10^5$ pfu/mouse of vcp2086 or the $2x10^7$ pfu/mouse of the ALVAC(2) parental vector. On day 14, the mice were boosted with $2x10^7$ pfu/mouse of vcp2086 or the ALVAC(2) parental vector. On day 15, spleen cells were isolated from five mice in each immunization group. On day 35, CTL were re-stimulated with peptides. On days 41, 50 and 55, ELISPOT assays were performed to detect IFN-γ producing T cells. Responses above control were observed for CEA-233 in studies MAD-69 and MAD-70. Responses above control were observed for p53-216 in study MAD-70.

CTL assays were also performed to detect cytotoxic T cells specific for CEA or p53. Cytotoxicity above control levels was observed following stimulation with CEA-233 or p53-216.

The data indicates that the multiantigen vector vcp2086 (ALVAC-CEA-p53-TRICOM) is capable of inducing anti-CEA and anti-p53 immune responses. It is shown that tolerance can be broken using ALVAC recombinants expressing CEA.

EXAMPLE 3

Modified Tumor Antigen KSA

A. Construction of Modified KSA

The tumor antigen KSA has been previously described (see, for example, Bjork, et al. J. Biol. Chem. 268:24232; Linnenbach, et al. Mol. and Cell. Biol. 13:1507; Szala, et al. PNAS 87:3542-3546; Balzar, et al. Journal of Molecular Medicine (1999), 77:699-712; and,

U.S. Pat. No. 5,348,887). A modified version of KSA was synthesized in order to increase the capacity of the antigen to generate an immune response by, for example, increasing the ability of KSA to bind MHC molecules. KSA may be modified by changing any of several amino acids to effect the desired change in the antigen. The sequences of the wild-type KSA (GenBank M33011; Szala, et al. PNAS 87:3542-3546) and KSA containing a particular modification utilized herein are aligned in **Figure 3** (sequence 1 represents M33011; sequence 2 represents the modified sequence; the modified sequences are indicated by an underline). In this manner, the T-cell epitope QLDPKFITSI (175-184) was converted to QLDPKFITSV. Synthesis of the modified KSA sequence is described below.

B. Expression Constructs

The cDNA clone in plasmid pRW971 encoding the GA733-2 carcinoma-associated antigen (KSA) was obtained from A. Linnenbach, The Wistar Institute, Philadelphia, PA. A XmaI-Spe I fragment containing the H6 promoter-KSA sequence was isolated from pRW971 and inserted into XmaI-SpeI sites on pBluescript to generate pBlu-KSA-1(R) (**Figure 4A**). To convert the codon ATT (Ile) at aa 184 of KSA to codon GTG (Val), the pBlu-KSA-1 was subjected to mutagenesis using a Stratagene kit and primers 8109 (CAAAATTTATCACGAGT(GTG)TTGTATGAGAATAATG) and 8110 (CATTATTCTCATACAA(CAC)ACTCGTGATAAATTTTG). The resulted plasmid mutant was designated pBlue-KSA-Val # 1 (**Figure 4A**). A XmaI-SpeI fragment was isolated from pBlue-KSA-Val #1 and inserted into the XmaI-SpeI sites on pT2255 generating pT2255-KSAV-1 (**Figure 4B**). A detailed plasmid map DNA sequence of pT2255-KSAV-1 are shown in **Figures 5A** and **B**, respectively.

The cDNA encoding LFA-3 was isolated at the National Cancer Institute by PCR amplification of Human Spleen Quick-Clone cDNA (Clontech Inc.) using the published sequence (Wallner et al. J. Exp. Med. 166:923-932, 1987). The cDNA encoding ICAM-1 was isolated at the National Cancer Institute by PCR amplification of cDNA reverse-transcribed from RNA from an Epstein-Barr Virus-transformed B cell line derived from a healthy male, using the published sequence (Staunton et al. Cell 52:925-933, 1988). The cDNA encoding B7.1 was isolated at the National Cancer Institute by PCR amplification of cDNA derived from RNA from the human Raji cell line (ATCC # CCL 86), using the published sequence (Chen et al. Cell 71:1093-1102, 1992).

As previously described elsewhere, vCP1468 (ALVAC(2)) was generated by insertion of the vaccinia virus E3L and K3L genes into the C6 site of parental ALVAC using the donor plasmid pMPC6H6K3E3. vCP2041 was generated by insertion of the LFA-3, ICAM-1 and B7.1 genes into the C3 sites of the recombinant ALVAC vCP1468 (ALVAC(2)) using the donor plasmid pALVAC.Tricom(C3) #33 (**Figure 6**). vCP2055 was generated by insertion of the KSA gene into the C5 sites of the recombinant ALVAC vCP2041 using the donor plasmid pT2255KSA(Val)LM (**Figure 6**). **Tables 2-4** further describe the arrangement of this expression vector.

Table 2. Authentic Gene Product(s)

Gene	Molecular Weight (kD)	Known Processing Events	Subcellular Localization
E3L	21.5; runs as 25	also a 20 kDa protein from internal initiation	nuclear
K3L	10	not relevant	not relevant
LFA-3	55-70	glycosylation	cell surface (transmembrane)
ICAM-1	90-110	glycosylation	cell surface (transmembrane)
B7.1	60	glycosylation	cell surface (transmembrane)
KSA	40	glycosylation	transmembrane

Table 3: Promoter(s)

Gene	Promoter
E3L	vaccinia E3L
K3L	vaccinia H6
LFA-3	vaccinia 30K
ICAM-1	vaccinia I3
B7.1	sE/L
KSA	vaccinia H6

Table 4: Donor Plasmids

Name	Size (bp)	Vector	Antibiotic Resitance Gene	Map Attached
pMPC6H6K3E3	7,400	pBS-SK	Amp	No
pALVAC.Tricom(C3) #33	10,470	pBS-SK	Amp	Yes
pT2255KSA(Val)LM	9,515	pBS-SK	Amp	Yes

CEF cells were infected with the expression vector using standard techniques. The modified KSA expressed in the CEF cells was analyzed by Western blot. The modified KSA is a glycoprotein with 314 amino acids. The protein expressed by ALVAC was shown to be 40 Kd on Western blot (data not shown). Thus, the modified KSA protein is expressed from the ALVAC expression vector.

It is also possible to incorporate the modified KSA coding sequence into an expression vector encoding other tumor antigens. For instance, it may be beneficial to insert the modified KSA sequence into ALVAC-CEA-p53-TRICOM to effectuate expression of CEA, p53, KSA, and the co-stimulatory components from a single vector.

EXAMPLE 4

Multi-Antigen Cancer Vaccine

The vectors described herein are useful for generating anti-cancer immune responses. The vectors are especially useful for generating anti-cancer immune responses where the tumor expresses multiple tumor antigens. For instance, a colorectal cancer may express CEA, p53 and KSA. In such a case, it may be useful to administer ALVAC-CEA-p53-TRICOM alone or in combination with the ALVAC vector vCP2055 to generate an anti-tumor immune response. The vector or vectors may be administered in separate pharmaceutically acceptable compositions or as a single pharmaceutically acceptable composition. Where multiple vectors are utilized, the vectors may be administered at a single site or at separate sites within the host. As such, an anti-tumor immune response is generated which decreases or halts tumor growth by the anti-tumor activity of immune cells such as cytotoxic T cells of the host.

While the present invention has been described in terms of the preferred embodiments, it is understood that variations and modifications will occur to those skilled in

the art. Therefore, it is intended that the appended claims cover all such equivalent variations that come within the scope of the invention as claimed.

CLAIMS

What is claimed is:

1. An expression vector useful for immunizing a host comprising nucleic acid sequences encoding modified KSA.
2. The expression vector of claim 1 wherein the vector is a plasmid or a viral vector.
3. The expression vector of claim 2 wherein the viral vector is selected from the group consisting of poxvirus, adenovirus, retrovirus, herpesvirus, and adeno-associated virus.
4. The expression vector of claim 3 wherein the viral vector is a poxvirus selected from the group consisting of vaccinia, NYVAC, avipox, canarypox, ALVAC, ALVAC(2), fowlpox, and TROVAC.
5. The expression vector of claim 4 wherein the viral vector is a poxvirus selected from the group consisting of NYVAC, ALVAC, and ALVAC(2).
6. The expression vector of claim 1 further comprising at least one additional tumor-associated antigen.
7. The expression vector of claim 6 wherein the vector is a plasmid or a viral vector.
8. The expression vector of claim 7 wherein the viral vector is selected from the group consisting of poxvirus, adenovirus, retrovirus, herpesvirus, and adeno-associated virus.
9. The expression vector of claim 8 wherein the viral vector is a poxvirus selected from the group consisting of vaccinia, MVA, NYVAC, avipox, canarypox, ALVAC, ALVAC(2), fowlpox, and TROVAC.
10. The expression vector of claim 9 wherein the viral vector is a poxvirus selected from the group consisting of NYVAC, ALVAC, and ALVAC(2).
11. The expression vector of claim 1 further comprising at least one nucleic sequence encoding an angiogenesis-associated antigen.
12. The expression vector of claim 11 wherein the vector is a plasmid or a viral vector.
13. The expression vector of claim 12 wherein the viral vector is selected from the group consisting of poxvirus, adenovirus, retrovirus, herpesvirus, and adeno-associated virus.
14. The expression vector of claim 13 wherein the viral vector is a poxvirus selected from the group consisting of vaccinia, MVA, NYVAC, avipox, canarypox, ALVAC, ALVAC(2), fowlpox, and TROVAC.
15. The expression vector of claim 14 wherein the viral vector is a poxvirus selected from the group consisting of NYVAC, ALVAC, and ALVAC(2).

16. The expression vector of claim 6 further comprising at least one nucleic sequence encoding an angiogenesis-associated antigen.
17. The expression vector of claim 16 wherein the vector is a plasmid or a viral vector.
18. The expression vector of claim 17 wherein the viral vector is selected from the group consisting of poxvirus, adenovirus, retrovirus, herpesvirus, and adeno-associated virus.
19. The expression vector of claim 17 wherein the viral vector is a poxvirus selected from the group consisting of vaccinia, MVA, NYVAC, avipox, canarypox, ALVAC, ALVAC(2), fowlpox, and TROVAC.
20. The poxvirus of claim 18 wherein the viral vector is a poxvirus selected from the group consisting of NYVAC, ALVAC, and ALVAC(2).
21. The expression vector of claim 1, 6, 11 or 16 further comprising at least one nucleic acid sequence encoding a co-stimulatory component.
22. The expression vector of claim 21 wherein the co-stimulatory component is selected from the group consisting of B7.1, LFA-3 and ICAM-1.
23. The expression vector of claim 22 or 23 wherein the vector is a plasmid or a viral vector.
24. The expression vector of claim 23 wherein the viral vector is selected from the group consisting of poxvirus, adenovirus, retrovirus, herpesvirus, and adeno-associated virus.
25. The expression vector of claim 24 wherein the viral vector is a poxvirus selected from the group consisting of vaccinia, MVA, NYVAC, avipox, canarypox, ALVAC, ALVAC(2), fowlpox, and TROVAC.
26. The poxvirus of claim 25 wherein the viral vector is a poxvirus selected from the group consisting of NYVAC, ALVAC, and ALVAC(2).
27. A composition comprising an expression vector in a pharmaceutically acceptable carrier, said vector comprising nucleic acid sequences encoding modified KSA.
28. The expression vector of claim 27 wherein the vector is a plasmid or a viral vector.
29. The expression vector of claim 28 wherein the viral vector is selected from the group consisting of poxvirus, adenovirus, retrovirus, herpesvirus, and adeno-associated virus.
30. The expression vector of claim 29 wherein the viral vector is a poxvirus selected from the group consisting of vaccinia, MVA, NYVAC, avipox, canarypox, ALVAC, ALVAC(2), fowlpox, and TROVAC.
31. The poxvirus of claim 30 wherein the viral vector is a poxvirus selected from the group consisting of NYVAC, ALVAC, and ALVAC(2).

32. A method for preventing or treating cancer comprising administering to a host an expression vector comprising nucleic acid sequences encoding modified KSA.

33. The expression vector of claim 32 wherein the vector is a plasmid or a viral vector.

34. The expression vector of claim 33 wherein the viral vector is selected from the group consisting of poxvirus, adenovirus, retrovirus, herpesvirus, and adeno-associated virus.

35. The expression vector of claim 34 wherein the viral vector is a poxvirus selected from the group consisting of vaccinia, MVA, NYVAC, avipox, canarypox, ALVAC, ALVAC(2), fowlpox, and TROVAC.

36. The poxvirus of claim 35 wherein the viral vector is a poxvirus selected from the group consisting of NYVAC, ALVAC, and ALVAC(2).

36. An isolated DNA molecule comprising the modified KSA coding sequence illustrated in Figure 3.

36. An isolated DNA molecule comprising a nucleotide sequence encoding modified KSA having the amino acid sequence shown in Figure 3.

37. An isolated DNA molecule comprising CEA, p53, and modified KSA coding sequences, the CEA sequence being CEA-CAP1-6D-1,2 as illustrated in Figure 2, the p53 sequence being the p53 sequence illustrated in Figure 1, and the modified KSA sequence being that shown in Figure 3.

FIGURE 1

Plasmid sequence of pNC5LSPCEAp53 (pMC30B5) for vCP2086

```
   1  GCCCTTT CGTCTCG CGCGTTT CGGTGAT GACGGTG AAAACCT CTGACAC ATGCAGC TCCCGGA GACGGTC
      CGGGAAA GCAGAGC GCGCAAA GCCACTA CTGCCAC TTTTGGA GACTGTG TACGTCG AGGGCCT CTGCCAG
  71  ACAGCTT GTCTGTA AGCGGAT GCCGGGA GCAGACA AGCCCGT CAGGGCG CGTCAGC GGGTGTT GGCGGGT
      TGTCGAA CAGACAT TCGCCTA CGGCCCT CGTCTGT TCGGGCA GTCCCGC GCAGTCG CCCACAA CCGCCCA
 141  GTCGGGG CTGGCTT AACTATG CGGCATC AGAGCAG ATTGTAC TGAGAGT GCACCAT ATGCGGT GTGAAAT
      CAGCCCC GACCGAA TTGATAC GCCGTAG TCTCGTC TAACATG ACTCTCA CGTGGTA TACGCCA CACTTTA
 211  ACCGCAC AGATGCG TAAGGAG AAAATAC CGCATCA GGCGCCA TTCGCCA TTCAGGC TGCGCAA CTGTTGG
      TGGCGTG TCTACGC ATTCCTC TTTTATG GCGTAGT CCGCGGT AAGCGGT AAGTCCG ACGCGTT GACAACC
 281  GAAGGGC GATCGGT GCGGGCC TCTTCGC TATTACG CCAGCTG GCGAAAG GGGGATG TGCTGCA AGGCGAT
      CTTCCCG CTAGCCA CGCCCGG AGAAGCG ATAATGC GGTCGAC CGCTTTC CCCCTAC ACGACGT TCCGCTA
 351  TAAGTTG GGTAACG CCAGGGT TTTCCCA GTCACGA CGTTGTA AAACGAC GGCCAGT GCCAAGC TTGGCTG
      ATTCAAC CCATTGC GGTCCCA AAAGGGT CAGTGCT GCAACAT TTTGCTG CCGGTCA CGGTTCG AACCGAC
      ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
                                                                              Left Arm
 421  CAGGTAT TCTAAAC TAGGAAT AGATGAA ATTATGT GCAAAGG AGATACC TTTAGAT ATGGATC TGATTTA
      GTCCATA AGATTTG ATCCTTA TCTACTT TAATACA CGTTTCC TCTATGG AAATCTA TACCTAG ACTAAAT
                                          Left Arm
 491  TTTGGTT TTTCATA ATCATAA TCTAACA ACATTTT CACTATA CTATACC TTCTTGC ACAAGTC GCCATTA
      AAACCAA AAAGTAT TAGTATT AGATTGT TGTAAAA GTGATAT GATATGG AAGAACG TGTTCAG CGGTAAT
                                          Left Arm
 561  GTAGTAT AGACTTA TACTTTG TAACCAT AGTATAC TTTAGCG CGTCATC TTCTTCA TCTAAAA CAGATTT
      CATCATA TCTGAAT ATGAAAC ATTGGTA TCATATG AAATCGC GCAGTAG AAGAAGT AGATTTT GTCTAAA
                                          Left Arm
 631  ACAACAA TAATCAT CGTCGTC ATCTTCA TCTTCAT TAAAGTT TTCATAT TCAATAA CTTTCTT TTCTAAA
      TGTTGTT ATTAGTA GCAGCAG TAGAAGT AGAAGTA ATTTCAA AAGTATA AGTTATT GAAAGAA AAGATTT
                                          Left Arm
 701  ACATCAT CTGAATC AATAAAC ATAGAAC GGTATAG AGCGTTA ATCTCCA TTGTAAA ATATACT AACGCGT
      TGTAGTA GACTTAG TTATTTG TATCTTG CCATATC TCGCAAT TAGAGGT AACATTT TATATGA TTGCGCA
                                          Left Arm
 771  TGCTCAT GATGTAC TTTTTTT CATTATT TAGAAAT TATGCAT TTTAGAT CTTTATA AGCGGCC GTGATTA
      ACGAGTA CTACATG AAAAAAA GTAATAA ATCTTTA ATACGTA AAATCTA GAAATAT TCGCCGG CACTAAT
                                                                          ~~~
                                        Left Arm
 841  ACTAGTC ATAAAAA CCCGGGA TCGATTC TAGACTC GAGATAA AAACTAT ATCAGAG CAACCCC AACCAGC
      TGATCAG TATTTTT GGGCCCT AGCTAAG ATCTGAG CTCTATT TTTGATA TAGTCTC GTTGGGG TTGGTCG
      ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
                                                                    .CEA
                                                         ***Ile LeuAla ValGly ValLeuVal·
 911  ACTCCAA TCATGAT GCCGACA GTGGCCC CAGCTGA GAGACCA GGAGAAG TTCCAGA TGCAGAG ACTGTGA
      TGAGGTT AGTACTA CGGCTGT CACCGGG GTCGACT CTCTGGT CCTCTTC AAGGTCT ACGTCTC TGACACT
                                          CEA
      ..GlyIle MetIle GlyValThr AlaGly AlaSer LeuGlyPro SerThr GlySer AlaSerVal ThrIle·
 981  TGCTCTT GACTATG GAATTAT TGCGGCC AGTAGCC AAGTTAG AGACAAA ACAGGCA TAGGTCC CGTTATT
      ACGAGAA CTGATAC CTTAATA ACGCCGG TCATCGG TTCAATC TCTGTTT TGTCCGT ATCCAGG GCAATAA
                                          CEA
      .SerLys ValIleSer AsnAsn ArgGly ThrAlaLeu AsnSer ValPhe CysAlaTyr ThrGly AsnAsn
1051  ATTTGGC GTGATTT TGGCGAT AAAGAGA ACTTGTG TGTGTTG CTGCGGT ATCCCAT TGATACG CCAAGAA
      TAAACCG CACTAAA ACCGCTA TTTCTCT TGAACAC ACACAAC GACGCCA TAGGGTA ACTATGC GGTTCTT
                                          CEA
      AsnProThr IleLys AlaIle PheLeuVal GlnThr HisGln GlnProIle GlyAsn IleArg TrpSerTyr·
1121  TACTGCG GGGATGG GTTAGAG GCCGAGT GGCAGGA GAGGTTG AGGTCCG CTCCCGA AAGGTAA GACGAGT
      ATGACGC CCCTACC CAATCTC CGGCTCA CCGTCCT CTCCAAC TCCAGGC GAGGGCT TTCCATT CTGCTCA
                                          CEA
      ..GlnPro SerPro AsnSerAla SerHis CysSer LeuAsnLeu AspAla GlySer LeuTyrSer SerAsp·
1191  CTGGGGG GGAAATG ATGGGGG TGTCCGG CCCATAG AGGACAT CCAGGGT GACTGGG TCACTGC GGTTTGC
      GACCCCC CCTTTAC TACCCCC ACAGGCC GGGTATC TCCTGTA GGTCCCA CTGACCC AGTGACG CCAAACG
                                          CEA
      .ProPro SerIleIle ProThr AspPro GlyTyrLeu ValAsp LeuThr ValProAsp SerArg AsnAla
1261  ACTCACT GAGTTCT GGATTCC ACATACA TAGGCTC TTGCGTC ATTTCTT GTGACAT TGAATAG AGTGAGG
      TGAGTGA CTCAAGA CCTAAGG TGTATGT ATCCGAG AACGCAG TAAAGAA CACTGTA ACTTATC TCACTCC
                                          CEA
      SerValSer AsnGln IleGly CysValTyr AlaArg AlaAsp AsnArgThr ValAsn PheLeu ThrLeuThr·
1331  GTCCTGT TGCCATT GGACAGC TGCAGCC TGGGACT GACTGGG AGGCTCT GACCATT TACCCAC CACAGGT
      CAGGACA ACGGTAA CCTGTCG ACGTCGG ACCCTGA CTGACCC TCCGAGA CTGGTAA ATGGGTG GTGTCCA
                                          CEA
      ..ArgAsn GlyAsn SerLeuGln LeuArg ProSer ValProLeu SerGln GlyAsn ValTrpTrp LeuTyr·
1401  AGGTTGT GTTCTGA GCCTCAG GTTCACA GGTGAAG GCCACAG CATCCTT GTCCTCC ACGGGTT TGGAGTT
      TCCAACA CAAGACT CGGAGTC CAAGTGT CCACTTC CGGTGTC GTAGGAA CAGGAGG TGCCCAA ACCTCAA
```

```
                                    CEA
       .ThrThr AsnGlnAla GluPro GluCys ThrPheAla ValAla AspLys AspGluVal ProLys SerAsn
1471   GTTGCTG GAGATGG AGGGCTT GGGCAGC TCCGCGG AAACAGT TATTGTT TTAACTG TAGTCCT GCTGTGA
       CAACGAC CTCTACC TCCCGAA CCCGTCG AGGCGCC TTTGTCA ATAACAA AATTGAC ATCAGGA CGACACT
                                    CEA
       AsnSerSer IleSer ProLys ProLeuGlu AlaSer ValThr IleThrLys ValThr ThrArg SerHisGly·
1541   CCACTGG CTGAGTT ATTGGCC TGGCAAG TATAGAG TCCGCTG TTCTTCT CAGTTAT GTTGCTT ATAAATA
       GGTGACC GACTCAA TAACCGG ACCGTTC ATATCTC AGGCGAC AAGAAGA GTCAATA CAACGAA TATTTAT
                                    CEA
       ..SerAla SerAsn AsnAlaGln CysThr TyrLeu GlySerAsn LysGlu ThrIle AsnSerIle PheLeu·
1611   ACTCTTG AGTATGC TGCTGAA TGTTTCC ATCAATC AGCCAGG AGTACTG TGCAGGG GGTTGG ATGCTGC
       TGAGAAC TCATACG ACGACTT ACAAAGG TAGTTAG TCGGTCC TCATGAC ACGTCCC CCCAACC TACGACG
                                    CEA
       .GluGln ThrHisGln GlnIle AsnGly AspIleLeu TrpSer TyrGln AlaProPro AsnSer AlaAla
1681   ATGGCAA GAAAGGC TCAAGTT CACGCCG GGACGGT AGTAGGT GTATGAT GGAGATA TAGTTGG GTCGTCT
       TACCGTT CTTTCCG AGTTCAA GTGCGGC CCTGCCA TCATCCA CATACTA CCTCTAT ATCAACC CAGCAGA
                                    CEA
       HisCysSer LeuSer LeuAsn ValGlyPro ArgTyr TyrThr TyrSerPro SerIle ThrPro AspAspPro·
1751   GGGCCAT ACAAAAC ATTAAGG ATAACAG GGTCGGA GTGATCA ACGGATA ATTCATT CTGAATG CCACACT
       CCCGGTA TGTTTTG TAATTCC TATTGTC CCAGCCT CACTAGT TGCCTAT TAAGTAA GACTTAC GGTGTGA
                                    CEA
       ..GlyTyr LeuVal AsnLeuIle ValPro AspSer HisAspVal SerLeu GluAsn GlnIleGly CysGlu·
1821   CATAAGG TCCTACA TCATTGC GAGTAAC GGACAGG AGTGTCA ATGTGCG GTTATCA TTAGACA ACTGCAA
       GTATTCC AGGATGT AGTAACG CTCATTG CCTGTCC TCACAGT TACACGC CAATAGT AATCTGT TGACGTT
                                    CEA
       .TyrPro GlyValAsp AsnArg ThrVal SerLeuLeu ThrLeu ThrArg AsnAspAsn SerLeu GlnLeu
1891   GCGTGGG CTAACCG GCAAACT TTGGTTA TTGACCC ACCATAA ATAAGTG GTATTTT GAATCTC TGGCTCA
       CGCACCC GATTGGC CGTTTGA AACCAAT AACTGGG TGGTATT TATTCAC CATAAAA CTTAGAG ACCGAGT
                                    CEA
       ArgProSer ValPro LeuSer GlnAsnAsn ValTrp TrpLeu TyrThrThr AsnGln IleGlu ProGluCys·
1961   CAAGTTA ATGCAAC TGCGTCC TCATCCT CAACTGG GTTAGAA TTGTTAC TAGTTAT GAATGGT TTTGGTG
       GTTCAAT TACGTTG ACGCAGG AGTAGGA GTTGACC CAATCTT AACAATG ATCAATA CTTACCA AAACCAC
                                    CEA
       ..ThrLeu AlaVal AlaAspGlu AspGlu ValPro AsnSerAsn AsnSer ThrIle PheProLys ProPro·
2031   GCTCATA CACGGTA ATCGTCG TCACGGT TGTGCGG TTGAGTC CGGTGTC GCTATTG TGAGCTT GGCACGT
       CGAGTAT GTGCCAT TAGCAGC AGTGCCA ACACGCC AACTCAG GCCACAG CGATAAC ACTCGAA CCGTGCA
                                    CEA
       .GluTyr ValThrIle ThrThr ValThr ThrArgAsn LeuGly ThrAsp SerAsnHis AlaGln CysThr
2101   GTAGGAT CCACTAT TGTTCAC GGTAATA TTGGGAA TGAACAG TTCCTGG GTGGACT GTTGGAA AGTGCCA
       CATCCTA GGTGATA ACAAGTG CCATTAT AACCCTT ACTTGTC AAGGACC CACCTGA CAACCTT TCACGGT
                                    CEA
       TyrSerGly SerAsn AsnVal ThrIleAsn ProIle PheLeu GluGlnThr SerGln GlnPhe ThrGlyAsn·
2171   TTGACAA ACCAGCT GTATTGG GCGGGAG GATTGCT AGCGGCA TGACAGC TCAGATT CAGATTT TCCCCTG
       AACTGTT TGGTCGA CATAACC CGCCCTC CTAACGA TCGCCGT ACTGTCG AGTCTAA GTCTAAA AGGGGAC
                                    CEA
       ..ValPhe TrpSer TyrGlnAla ProPro AsnSer AlaAlaHis CysSer LeuAsn LeuAsnGlu GlySer·
2241   ATCTATA GCTTGTG TTTAGAG GGCTGAT TGTAGGA GCATCGG GTCCGTA AAGCACG TTGAGAA TCACTGA
       TAGATAT CGAACAC AAATCTC CCGACTA ACATCCT CGTAGCC CAGGCAT TTCGTGC AACTCTT AGTGACT
                                    CEA
       .ArgTyr SerThrAsn LeuPro SerIle ThrProAla AspPro GlyTyr LeuValAsn LeuIle ValSer
2311   ATCAGAC CTCCTGG CGCTGAC TGGATTT TGGGTTT CGCATTT GTAGCTT GCTGTGT CGTTCCT GGTCACG
       TAGTCTG GAGGACC GCGACTG ACCTAAA ACCCAAA GCGTAAA CATCGAA CGACACA GCAAGGA CCAGTGC
                                    CEA
       AspSerArg ArgAla SerVal ProAsnGln ThrGlu CysLys TyrSerAla ThrAsp AsnArg ThrValAsn·
2381   TTAAACA GGGTCAG AGTTCTA TTTCCGT TGCTGAG TTGGAGT CTAGGGG ACACAGG CAGGGAC TGGTTGT
       AATTTGT CCCAGTC TCAAGAT AAAGGCA ACGACTC AACCTCA GATCCCC TGTGTCC GTCCCTG ACCAACA
                                    CEA
       ..PheLeu ThrLeu ThrArgAsn GlyAsn SerLeu GlnLeuArg ProSer ValPro LeuSerGln AsnAsn·
2451   TCACCCA CCAGAGA TATGTTG CGTCTTG AGTTTCG GGCTCGC ATGTAAA AGCGACG GCATCTT TGTCTTC
       AGTGGGT GGTCTCT ATACAAC GCAGAAC TCAAAGC CCGAGCG TACATTT TCGCTGC CGTAGAA ACAGAAG
                                    CEA
       .ValTrp TrpLeuTyr ThrAla AspGln ThrGluPro GluCys ThrPhe AlaValAla AspLys AspGlu
2521   GACAGGC TTACTAT TATTGGA GCTAATA GAAGGCT TAGGGAG TTCCGGG TATACCC GGAACTG GCCAGTT
       CTGTCCG AATGATA ATAACCT CGATTAT CTTCCGA ATCCCTC AAGGCCC ATATGGG CCTTGAC CGGTCAA
                                    CEA
       ValProLys SerAsn AsnSer SerIleSer ProLys ProLeu GluProTyr ValArg PheGln GlyThrAla·
2591   GCTTCTT CATTCAC AAGATCT GACTTTA TGACGTG TAGGGTG TAGAATC CTGTGTC ATTCTGG ATGATGT
       CGAAGAA GTAAGTG TTCTAGA CTGAAAT ACTGCAC ATCCCAC ATCTTAG GACACAG TAAGACC TACTACA
                                    CEA
       ..GluGlu AsnVal LeuAspSer LysIle ValHis LeuThrTyr PheGly ThrAsp AsnGlnIle IleAsn·
2661   TCTGGAT CAGCAGG GATGCAT TGGGGTA TATTATC TCTCGAC CACTGTA TGCGGGC CCTGGGG TAGCTTG
       AGACCTA GTCGTCC CTACGTA ACCCCAT ATAATAG AGAGCTG GTGACAT ACGCCCG GGACCCC ATCGAAC
                                    CEA
```

```
        .GlnIle LeuLeuSer AlaAsn ProTyr IleIleGlu ArgGly SerTyr AlaProGly ProThr AlaGln
2731    TTGAGTT CCTATTA CATATCC TATAATT TGACGGT TGCCATC CACTCTT TCACCTT TGTACCA GCTGTAG
        AACTCAA GGATAAT GTATAGG ATATTAA ACTGCCA ACGGTAG GTGAGAA AGTGGAA ACATGGT CGACATC
                                          CEA
        GlnThrGly IleVal TyrGly IleIleGln ArgAsn GlyAsp ValArgGlu GlyLys TyrTrp SerTyrGly·
2801    CCAAAAA GATGCTG GGGCAGA TTGTGGA CAAGTAG AAGCACC TCCTTCC CCTCTGC GACATTG AACGGCG
        GGTTTTT CTACGAC CCCGTCT AACACCT GTTCATC TTCGTGG AGGAAGG GGAGACG CTGTAAC TTGCCGC
                                          CEA
        ..PheLeu HisGln ProLeuAsn HisVal LeuLeu LeuValGlu LysGly GluAla ValAsnPhe ProThr·
2871    TGGATTC AATAGTG AGCTTGG CAGTGGT GGGCGGG TTCCAGA AGGTTAG AAGTGAG GCTGTGA GCAGGAG
        ACCTAAG TTATCAC TCGAACC GTCACCA CCCGCCC AAGGTCT TCCAATC TTCACTC CGACACT CGTCCTC
                                          CEA
        .SerGlu IleThrLeu LysAla ThrThr ProProAsn TrpPhe ThrLeu LeuSerAla ThrLeu LeuLeu
2941    CCTCTGC CAGGGGA TGCACCA TCTGTGG GGAGGGG CCGAGGG AGACTCC ATTATTT ATATTCC AAAAAAA
        GGAGACG GTCCCCT ACGTGGT AGACACC CCTCCCC GGCTCCC TCTGAGG TAATAAA TATAAGG TTTTTTT
        ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
                                                                    E/L Promoter
                                                               ~~~~~~~~~~~~~~~~~~~~~
                                CEA
        ArgGlnTrp ProIle CysTrp ArgHisPro ProAla SerPro SerGluMet
                                                                     H6 promoter
                                                                   ~~~~~~~~~~~~
3011    AAAAATA AAATTTC AATTTTT GTCGACC TGCAGCT CGACGGA TCCCCCC GGGTTCT TTATTCT ATACTTA
        TTTTTAT TTTAAAG TTAAAAA CAGCTGG ACGTCGA GCTGCCT AGGGGGG CCCAAGA AATAAGA TATGAAT
                                   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
            E/L Promoter
                                       H6 promoter
        ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
3081    AAAAGTG AAAATAA ATACAAA GGTTCTT GAGGGTT GTGTTAA ATTGAAA GCGAGAA ATAATCA TAAATTA
        TTTTCAC TTTTATT TATGTTT CCAAGAA CTCCCAA CACAATT TAACTTT CGCTCTT TATTAGT ATTTAAT
                                                                  p53
                                                         ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
                     H6 promoter
        ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
                                                         MetGlu GluProGln SerAsp ProSer ValGluPro·
3151    TTTCATT ATCGCGA TATCCGT TAAGTTT GTATCGT AATGGAG GAGCCGC AGTCAGA TCCTAGC GTCGAGC
        AAAGTAA TAGCGCT ATAGGCA ATTCAAA CATAGCA TTACCTC CTCGGCG TCAGTCT AGGATCG CAGCTCG
                                          p53
        ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
        ..ProLeu SerGln GluThrPhe SerAsp LeuTrp LysLeuLeu ProGlu AsnAsn ValLeuSer ProLeu·
3221    CCCCTCT GAGTCAG GAAACAT TTTCAGA CCTATGG AAACTAC TTCCTGA AAACAAC GTTCTGT CCCCCTT
        GGGGAGA CTCAGTC CTTTGTA AAAGTCT GGATACC TTTGATG AAGGACT TTTGTTG CAAGACA GGGGGAA
                                          p53
        ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
        .ProSer GlnAlaMet AspAsp LeuMet LeuSerPro AspAsp IleGlu GlnTrpPhe ThrGlu AspPro
3291    GCCGTCC CAAGCAA TGGATGA TTTGATG CTGTCCC CGGACGA TATTGAA CAATGGT TCACTGA AGACCCA
        CGGCAGG GTTCGTT ACCTACT AAACTAC GACAGGG GCCTGCT ATAACTT GTTACCA AGTGACT TCTGGGT
                                          p53
        ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
        GlyProAsp GluAla ProArg MetProGlu AlaAla ProPro ValAlaPro AlaPro AlaAla ProThrPro·
3361    GGTCCAG ATGAAGC TCCCAGA ATGCCAG AGGCTGC TCCCCCC GTGGCCC CTGCACC AGCAGCT CCTACAC
        CCAGGTC TACTTCG AGGGTCT TACGGTC TCCGACG AGGGGGG CACCGGG GACGTGG TCGTCGA GGATGTG
                                          p53
        ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
        ..AlaAla ProAla ProAlaPro SerTrp ProLeu SerSerSer ValPro SerGln LysThrTyr GlnGly·
3431    CGGCGGC CCCTGCA CCAGCCC CCTCCTG GCCCCTG TCATCTT CTGTCCC TTCCCAG AAAACCT ACCAGGG
        GCCGCCG GGGACGT GGTCGGG GGAGGAC CGGGGAC AGTAGAA GACAGGG AAGGGTC TTTTGGA TGGTCCC
                                          p53
        ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
        .SerTyr GlyPheArg LeuGly PheLeu HisSerGly ThrAla LysSer ValThrCys ThrTyr SerPro
3501    CAGCTAC GGTTTCC GTCTGGG CTTCTTG CATTCTG GGACAGC CAAGTCT GTGACTT GCACGTA CTCCCCT
        GTCGATG CCAAAGG CAGACCC GAAGAAC GTAAGAC CCTGTCG GTTCAGA CACTGAA CGTGCAT GAGGGGA
                                          p53
        ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
        AlaLeuAsn LysMet PheCys GlnLeuAla LysThr CysPro ValGlnLeu TrpVal AspSer ThrProPro·
3571    GCCCTCA ACAAGAT GTTTTGC CAACTGG CCAAGAC CTGCCCT GTGCAGC TGTGGGT TGATTCC ACACCCC
        CGGGAGT TGTTCTA CAAAACG GTTGACC GGTTCTG GACGGGA CACGTCG ACACCCA ACTAAGG TGTGGGG
                                          p53
        ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
        ..ProGly ThrArg ValArgAla MetAla IleTyr LysGlnSer GlnHis MetThr GluValVal ArgArg·
3641    CGCCCGG CACCCGC GTCCGCG CCATGGC CATCTAC AAGCAGT CACAGCA CATGACG GAGGTTG TGAGGCG
        GCGGGCC GTGGGCG CAGGCGC GGTACCG GTAGATG TTCGTCA GTGTCGT GTACTGC CTCCAAC ACTCCGC
                                          p53
```

```
        ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
        .CysPro HisHisGlu ArgCys SerAsp SerAspGly LeuAla ProPro GlnHisLeu IleArg ValGlu
3711    CTGCCCC CACCATG AGCGCTG CTCAGAT AGCGATG GTCTGGC CCCTCCT CAGCATC TTATCCG AGTGGAA
        GACGGGG GTGGTAC TCGCGAC GAGTCTA TCGCTAC CAGACCG GGGAGGA GTCGTAG AATAGGC TCACCTT
                                          p53
        ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
        GlyAsnLeu ArgVal GluTyr LeuAspAsp ArgAsn ThrPhe ArgHisSer ValVal ValPro TyrGluPro·
3781    GGAAATT TGCGTGT GGAGTAT TTGGATG ACAGAAA CACTTTT CGACATA GTGTGGT GGTGCCC TATGAGC
        CCTTTAA ACGCACA CCTCATA AACCTAC TGTCTTT GTGAAAA GCTGTAT CACACCA CCACGGG ATACTCG
                                          p53
        ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
        ..ProGlu ValGly SerAspCys ThrThr IleHis TyrAsnTyr MetCys AsnSer SerCysMet GlyGly·
3851    CGCCTGA GGTTGGC TCTGACT GTACCAC CATCCAC TACAACT ACATGTG TAACAGT TCCTGCA TGGGCGG
        GCGGACT CCAACCG AGACTGA CATGGTG GTAGGTG ATGTTGA TGTACAC ATTGTCA AGGACGT ACCCGCC
                                          p53
        ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
        .MetAsn ArgArgPro IleLeu ThrIle IleThrLeu GluAsp SerSer GlyAsnLeu LeuGly ArgAsn
3921    CATGAAC CGGAGGC CCATCCT CACCATC ATCACAC TGGAAGA CTCCAGT GGTAATC TACTGGG ACGGAAC
        GTACTTG GCCTCCG GGTAGGA GTGGTAG TAGTGTG ACCTTCT GAGGTCA CCATTAG ATGACCC TGCCTTG
                                          p53
        ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
        SerPheGlu ValArg ValCys AlaCysPro GlyArg AspArg ArgThrGlu GluGlu AsnLeu ArgLysLys·
3991    AGCTTTG AGGTGCG TGTTTGT GCCTGTC CTGGGAG AGACCGG CGCACAG AGGAAGA GAATCTC CGCAAGA
        TCGAAAC TCCACGC ACAAACA CGGACAG GACCCTC TCTGGCC GCGTGTC TCCTTCT CTTAGAG GCGTTCT
                                          p53
        ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
        ..GlyGlu ProHis HisGluLeu ProPro GlySer ThrLysArg AlaLeu ProAsn AsnThrSer SerSer·
4061    AAGGGGA GCCTCAC CACGAGC TGCCCCC AGGGAGC ACTAAGC GAGCACT GCCCAAC AACACCA GCTCCTC
        TTCCCCT CGGAGTG GTGCTCG ACGGGGG TCCCTCG TGATTCG CTCGTGA CGGGTTG TTGTGGT CGAGGAG
                                          p53
        ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
        .ProGln ProLysLys LysPro LeuAsp GlyGluTyr PheThr LeuGln IleArgGly ArgGlu ArgPhe
4131    TCCCCAG CCAAAGA AGAAACC ACTGGAT GGAGAAT ATTTCAC CCTTCAG ATCCGTG GGCGTGA GCGCTTC
        AGGGGTC GGTTTCT TCTTTGG TGACCTA CCTCTTA TAAAGTG GGAAGTC TAGGCAC CCGCACT CGCGAAG
                                          p53
        ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
        GluMetPhe ArgGlu LeuAsn GluAlaLeu GluLeu LysAsp AlaGlnAla GlyLys GluPro GlyGlySer·
4201    GAGATGT TCCGAGA GCTGAAT GAGGCCT TGGAACT CAAGGAT GCCCAGG CTGGGAA GGAGCCA GGGGGGA
        CTCTACA AGGCTCT CGACTTA CTCCGGA ACCTTGA GTTCCTA CGGGTCC GACCCTT CCTCGGT CCCCCCT
                                          p53
        ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
        ..ArgAla HisSer SerHisLeu LysSer LysLys GlyGlnSer ThrSer ArgHis LysLysLeu MetPhe·
4271    GCAGGGC TCACTCC AGCCACC TGAAGTC CAAAAAG GGTCAGT CTACCTC CCGCCAT AAAAAAC TCATGTT
        CGTCCCG AGTGAGG TCGGTGG ACTTCAG GTTTTTC CCAGTCA GATGGAG GGCGGTA TTTTTTG AGTACAA
                    p53
        ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
        .LysThr GluGlyPro AspSer Asp***
4341    CAAGACA GAAGGGC CTGACTC AGACTGA ACGCGTT TTTTATC CCGGGCT CGAGGGT ACCGGAT CCTTTTT
        GTTCTGT CTTCCCG GACTGAG TCTGACT TGCGCAA AAAATAG GGCCCGA GCTCCCA TGGCCTA GGAAAAA
4411    ATAGCTA ATTAGTC ACGTACC TTTGAGA GTACCAC TTCAGCT ACCTCTT TTGTGTC TCAGAGT AACTTTC
        TATCGAT TAATCAG TGCATGG AAACTCT CATGGTG AAGTCGA TGGAGAA AACACAG AGTCTCA TTGAAAG
        ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
                                                                       Right Arm
4481    TTTAATC AATTCCA AAACAGT ATATGAT TTTCCAT TTCTTTC AAAGATG TAGTTTA CATCTGC TCCTTTG
        AAATTAG TTAAGGT TTTGTCA TATACTA AAAGGTA AAGAAAG TTTCTAC ATCAAAT GTAGACG AGGAAAC
                                      Right Arm
4551    TTGAAAA GTAGCCT GAGCACT TCTTTTC TACCATG AATTACA GCTGGCA AGATCAA TTTTTCC CAGTTCT
        AACTTTT CATCGGA CTCGTGA AGAAAAG ATGGTAC TTAATGT CGACCGT TCTAGTT AAAAAGG GTCAAGA
                                      Right Arm
4621    GGACATT TTATTTT TTTTAAG TAGTGTG CTACATA TTTCAAT ATTTCCA GATTGTA CAGCGAT CATTAAA
        CCTGTAA AATAAAA AAAATTC ATCACAC GATGTAT AAAGTTA TAAAGGT CTAACAT GTCGCTA GTAATTT
                                      Right Arm
4691    GGAGTAC GTCCCAT GTTATCC AGCAAGT CAGTATC AGCACCT TTGTTCA ATAGAAG TTTAACC ATTGTTA
        CCTCATG CAGGGTA CAATAGG TCGTTCA GTCATAG TCGTGGA AACAAGT TATCTTC AAATTGG TAACAAT
                                      Right Arm
4761    AATTTTT ATTTGAT ACGGCTA TATGTAG AGGAGTT AACCGAT CCGTGTT TGAAATA TCTACAT CCGCCGA
        TTAAAAA TAAACTA TGCCGAT ATACATC TCCTCAA TTGGCTA GGCACAA ACTTTAT AGATGTA GGCGGCT
                                      Right Arm
4831    ATGAGCC AATAGAA GTTTAAC CAAATTA ACTTTGT TAAGGTA AGCTGCC AAACACA AAGGAGT AAAGCCT
        TACTCGG TTATCTT CAAATTG GTTTAAT TGAAACA ATTCCAT TCGACGG TTTGTGT TTCCTCA TTTCGGA
                                      Right Arm
4901    CCGCTGT AAAGAAC ATTGTTT ACATAGT TATTCTT CAACAGA TCTTTCA CTATTTT GTAGTCG TCTCTCA
        GGCGACA TTTCTTG TAACAAA TGTATCA ATAAGAA GTTGTCT AGAAAGT GATAAAA CATCAGC AGAGAGT
```

```
                              Right Arm
4971   ACACCGC ATCATGC AGACAAG AAGTTGT GCATTCA GTAACTA CAGGTTT AGCTCCA TACCTCA TCAAGAT
       TGTGGCG TAGTACG TCTGTTC TTCAACA CGTAAGT CATTGAT GTCCAAA TCGAGGT ATGGAGT AGTTCTA
                              Right Arm
5041   TTTTATA GCCTCGG TATTCTT GAACATT ACAGCCA TTTCAAG AGGAGAT TGTAGAG TACCATA TTCCGTG
       AAAATAT CGGAGCC ATAAGAA CTTGTAA TGTCGGT AAAGTTC TCCTCTA ACATCTC ATGGTAT AAGGCAC
                              Right Arm
5111   TTAGGGT CGAATCC ATTGTCC AAAAACC TATTTAG AGATGCA TTGTCAT TATCCAT GATAGCC TCACAGA
       AATCCCA GCTTAGG TAACAGG TTTTTGG ATAAATC TCTACGT AACAGTA ATAGGTA CTATCGG AGTGTCT
                              Right Arm
5181   CGTATAT GTAAGCC ATCTTGA ATGTATA ATTTTGT TGTTTTC AACAACC GCTCGTG AACAGCT TCTATAC
       GCATATA CATTCGG TAGAACT TACATAT TAAAACA ACAAAAG TTGTTGG CGAGCAC TTGTCGA AGATATG
                              Right Arm
5251   TTTTTCA TTTTCTT CATGATT AATATAG TTTACGG AATATAA GTATACA AAAAGTT TATAGTA ATCTCAT
       AAAAAGT AAAAGAA GTACTAA TTATATC AAATGCC TTATATT CATATGT TTTTCAA ATATCAT TAGAGTA
                              Right Arm
5321   AATATCT GAAACAC ATACATA AAACATG GAAGAAT TACACGA TGTCGTT GAGATAA ATGGCTT TTTATTG
       TTATAGA CTTTGTG TATGTAT TTTGTAC CTTCTTA ATGTGCT ACAGCAA CTCTATT TACCGAA AAATAAC
                              Right Arm
5391   TCATAGT TTACAAA TTCGCAG TAATCTT CATCTTT TACGAAT ATTGCAG AATCTGT TTTATCC AACCAGT
       AGTATCA AATGTTT AAGCGTC ATTAGAA GTAGAAA ATGCTTA TAACGTC TTAGACA AAATAGG TTGGTCA
                              Right Arm
5461   GATTTTT GTATAAT ATAACTG GTATCCT ATCTTCC GATAGAA TGCTGTT ATTTAAC ATTTTTG CACCTAT
       CTAAAAA CATATTA TATTGAC CATAGGA TAGAAGG CTATCTT ACGACAA TAAATTG TAAAAAC GTGGATA
                              Right Arm
5531   TAAGTTA CATCTGT CAAATCC ATCTTTC CAACTGA CTTTATG TAACGAT GCGAAAT AGCATTT ATCACTA
       ATTCAAT GTAGACA GTTTAGG TAGAAAG GTTGACT GAAATAC ATTGCTA CGCTTTA TCGTAAA TAGTGAT
                              Right Arm
5601   TGTCGTA CCCAATT ATCATGA CAAGATT CTCTTAA ATACGTA ATCTTAT TATCTCT TGCATAT TCGTAAT
       ACAGCAT GGGTTAA TAGTACT GTTCTAA GAGAATT TATGCAT TAGAATA ATAGAGA ACGTATA AGCATTA
                              Right Arm
5671   AGTAATT GTAAAGA GTATACG ATAACAG TATAGAT ATACACG TGATATA AATATTT AACCCCA TTCCTGA
       TCATTAA CATTTCT CATATGC TATTGTC ATATCTA TATGTGC ACTATAT TTATAAA TTGGGGT AAGGACT
                              Right Arm
5741   GTAAAAT AATTACG ATATTAC ATTTCCT TTTATTA TTTTTAT GTTTTAG TTATTTG TTAGGTT ATACAAA
       CATTTTA TTAATGC TATAATG TAAAGGA AAATAAT AAAAATA CAAAATC AATAAAC AATCCAA TATGTTT
                              Right Arm
5811   AATTATG TTTATTT GTGTATA TTTAAAG CGTCGTT AAGAATA AGCTTAG TTAACAT ATTATCG CTTAGGT
       TTAATAC AAATAAA CACATAT AAATTTC GCAGCAA TTCTTAT TCGAATC AATTGTA TAATAGC GAATCCA
                              Right Arm
5881   TTTGTAG TATTTGA ATCCTTT CTTTAAA TGGATTA TTTTTCC AATGCAT ATTTATA GCTTCAT CCAAAGT
       AAACATC ATAAACT TAGGAAA GAAATTT ACCTAAT AAAAAGG TTACGTA TAAATAT CGAAGTA GGTTTCA
                              Right Arm
5951   ATAACAT TTAACAT TCAGAAT TGCGGCC GCAATTC AATTCGT AATCATG GTCATAG CTGTTTC CTGTGTG
       TATTGTA AATTGTA AGTCTTA ACGCCGG CGTTAAG TTAAGCA TTAGTAC CAGTATC GACAAAG GACACAC
                                      ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

               Right Arm
6021   AAATTGT TATCCGC TCACAAT TCCACAC AACATAC GAGCCGG AAGCATA AAGTGTA AAGCCTG GGGTGCC
       TTTAACA ATAGGCG AGTGTTA AGGTGTG TTGTATG CTCGGCC TTCGTAT TTCACAT TTCGGAC CCCACGG
6091   TAATGAG TGAGCTA ACTCACA TTAATTG CGTTGCG CTCACTG CCCGCTT TCCAGTC GGGAAAC CTGTCGT
       ATTACTC ACTCGAT TGAGTGT AATTAAC GCAACGC GAGTGAC GGGCGAA AGGTCAG CCCTTTG GACAGCA
6161   GCCAGCT GCATTAA TGAATCG GCCAACG CGCGGGG AGAGGCG GTTTGCG TATTGGG CGCTCTT CCGCTTC
       CGGTCGA CGTAATT ACTTAGC CGGTTGC GCGCCCC TCTCCGC CAAACGC ATAACCC GCGAGAA GGCGAAG
6231   CTCGCTC ACTGACT CGCTGCG CTCGGTC GTTCGGC TGCGGCG AGCGGTA TCAGCTC ACTCAAA GGCGGTA
       GAGCGAG TGACTGA GCGACGC GAGCCAG CAAGCCG ACGCCGC TCGCCAT AGTCGAG TGAGTTT CCGCCAT
6301   ATACGGT TATCCAC AGAATCA GGGGATA ACGCAGG AAAGAAC ATGTGAG CAAAAGG CCAGCAA AAGGCCA
       TATGCCA ATAGGTG TCTTAGT CCCCTAT TGCGTCC TTTCTTG TACACTC GTTTTCC GGTCGTT TTCCGGT
6371   GGAACCG TAAAAAG GCCGCGT TGCTGGC GTTTTTC CATAGGC TCCGCCC CCCTGAC GAGCATC ACAAAAA
       CCTTGGC ATTTTTC CGGCGCA ACGACCG CAAAAAG GTATCCG AGGCGGG GGGACTG CTCGTAG TGTTTTT
6441   TCGACGC TCAAGTC AGAGGTG GCGAAAC CCGACAG GACTATA AAGATAC CAGGCGT TTCCCCC TGGAAGC
       AGCTGCG AGTTCAG TCTCCAC CGCTTTG GGCTGTC CTGATAT TTCTATG GTCCGCA AAGGGGG ACCTTCG
6511   TCCCTCG TGCGCTC TCCTGTT CCGACCC TGCCGCT TACCGGA TACCTGT CCGCCTT TCTCCCT TCGGGAA
       AGGGAGC ACGCGAG AGGACAA GGCTGGG ACGGCGA ATGGCCT ATGGACA GGCGGAA AGAGGGA AGCCCTT
6581   GCGTGGC GCTTTCT CATAGCT CACGCTG TAGGTAT CTCAGTT CGGTGTA GGTCGTT CGCTCCA AGCTGGG
       CGCACCG CGAAAGA GTATCGA GTGCGAC ATCCATA GAGTCAA GCCACAT CCAGCAA GCGAGGT TCGACCC
6651   CTGTGTG CACGAAC CCCCCGT TCAGCCC GACCGCT GCGCCTT ATCCGGT AACTATC GTCTTGA GTCCAAC
       GACACAC GTGCTTG GGGGGCA AGTCGGG CTGGCGA CGCGGAA TAGGCCA TTGATAG CAGAACT CAGGTTG
6721   CCGGTAA GACACGA CTTATCG CCACTGG CAGCAGC CACTGGT AACAGGA TTAGCAG AGCGAGG TATGTAG
       GGCCATT CTGTGCT GAATAGC GGTGACC GTCGTCG GTGACCA TTGTCCT AATCGTC TCGCTCC ATACATC
6791   GCGGTGC TACAGAG TTCTTGA AGTGGTG GCCTAAC TACGGCT ACACTAG AAGGACA GTATTTG GTATCTG
       CGCCACG ATGTCTC AAGAACT TCACCAC CGGATTG ATGCCGA TGTGATC TTCCTGT CATAAAC CATAGAC
6861   CGCTCTG CTGAAGC CAGTTAC CTTCGGA AAAAGAG TTGGTAG CTCTTGA TCCGGCA AACAAAC CACCGCT
       GCGAGAC GACTTCG GTCAATG GAAGCCT TTTTCTC AACCATC GAGAACT AGGCCGT TTGTTTG GTGGCGA
```

```
6931   GGTAGCG GTGGTTT TTTTGTT TGCAAGC AGCAGAT TACGCGC AGAAAAA AAGGATC TCAAGAA GATCCTT
       CCATCGC CACCAAA AAAACAA ACGTTCG TCGTCTA ATGCGCG TCTTTTT TTCCTAG AGTTCTT CTAGGAA
7001   TGATCTT TTCTACG GGGTCTG ACGCTCA GTGGAAC GAAAACT CACGTTA AGGGATT TTGGTCA TGAGATT
       ACTAGAA AAGATGC CCCAGAC TGCGAGT CACCTTG CTTTTGA GTGCAAT TCCCTAA AACCAGT ACTCTAA
7071   ATCAAAA AGGATCT TCACCTA GATCCTT TTAAATT AAAAATG AAGTTTT AAATCAA TCTAAAG TATATAT
       TAGTTTT TCCTAGA AGTGGAT CTAGGAA AATTTAA TTTTTAC TTCAAAA TTTAGTT AGATTTC ATATATA
7141   GAGTAAA CTTGGTC TGACAGT TACCAAT GCTTAAT CAGTGAG GCACCTA TCTCAGC GATCTGT CTATTTC
       CTCATTT GAACCAG ACTGTCA ATGGTTA CGAATTA GTCACTC CGTGGAT AGAGTCG CTAGACA GATAAAG
       ~~~~~~~~~~~~~~~~~~~~~~
                                           Amp resistance gene
7211   GTTCATC CATAGTT GCCTGAC TCCCCGT CGTGTAG ATAACTA CGATACG GGAGGGC TTACCAT CTGGCCC
       CAAGTAG GTATCAA CGGACTG AGGGGCA GCACATC TATTGAT GCTATGC CCTCCCG AATGGTA GACCGGG
                                 Amp resistance gene
7281   CAGTGCT GCAATGA TACCGCG AGACCCA CGCTCAC CGGCTCC AGATTTA TCAGCAA TAAACCA GCCAGCC
       GTCACGA CGTTACT ATGGCGC TCTGGGT GCGAGTG GCCGAGG TCTAAAT AGTCGTT ATTTGGT CGGTCGG
                                 Amp resistance gene
7351   GGAAGGG CCGAGCG CAGAAGT GGTCCTG CAACTTT ATCCGCC TCCATCC AGTCTAT TAATTGT TGCCGGG
       CCTTCCC GGCTCGC GTCTTCA CCAGGAC GTTGAAA TAGGCGG AGGTAGG TCAGATA ATTAACA ACGGCCC
                                 Amp resistance gene
7421   AAGCTAG AGTAAGT AGTTCGC CAGTTAA TAGTTTG CGCAACG TTGTTGC CATTGCT ACAGGCA TCGTGGT
       TTCGATC TCATTCA TCAAGCG GTCAATT ATCAAAC GCGTTGC AACAACG GTAACGA TGTCCGT AGCACCA
                                 Amp resistance gene
7491   GTCACGC TCGTCGT TTGGTAT GGCTTCA TTCAGCT CCGGTTC CCAACGA TCAAGGC GAGTTAC ATGATCC
       CAGTGCG AGCAGCA AACCATA CCGAAGT AAGTCGA GGCCAAG GGTTGCT AGTTCCG CTCAATG TACTAGG
                                 Amp resistance gene
7561   CCCATGT TGTGCAA AAAAGCG GTTAGCT CCTTCGG TCCTCCG ATCGTTG TCAGAAG TAAGTTG GCCGCAG
       GGGTACA ACACGTT TTTTCGC CAATCGA GGAAGCC AGGAGGC TAGCAAC AGTCTTC ATTCAAC CGGCGTC
                                 Amp resistance gene
7631   TGTTATC ACTCATG GTTATGG CAGCACT GCATAAT TCTCTTA CTGTCAT GCCATCC GTAAGAT GCTTTTC
       ACAATAG TGAGTAC CAATACC GTCGTGA CGTATTA AGAGAAT GACAGTA CGGTAGG CATTCTA CGAAAAG
                                 Amp resistance gene
7701   TGTGACT GGTGAGT ACTCAAC CAAGTCA TTCTGAG AATAGTG TATGCGG CGACCGA GTTGCTC TTGCCCG
       ACACTGA CCACTCA TGAGTTG GTTCAGT AAGACTC TTATCAC ATACGCC GCTGGCT CAACGAG AACGGGC
                                 Amp resistance gene
7771   GCGTCAA TACGGGA TAATACC GCGCCAC ATAGCAG AACTTTA AAAGTGC TCATCAT TGGAAAA CGTTCTT
       CGCAGTT ATGCCCT ATTATGG CGCGGTG TATCGTC TTGAAAT TTTCACG AGTAGTA ACCTTTT GCAAGAA
                                 Amp resistance gene
7841   CGGGGCG AAAACTC TCAAGGA TCTTACC GCTGTTG AGATCCA GTTCGAT GTAACCC ACTCGTG CACCCAA
       GCCCCGC TTTTGAG AGTTCCT AGAATGG CGACAAC TCTAGGT CAAGCTA CATTGGG TGAGCAC GTGGGTT
                                 Amp resistance gene
7911   CTGATCT TCAGCAT CTTTTAC TTTCACC AGCGTTT CTGGGTG AGCAAAA ACAGGAA GGCAAAA TGCCGCA
       GACTAGA AGTCGTA GAAAATG AAAGTGG TCGCAAA GACCCAC TCGTTTT TGTCCTT CCGTTTT ACGGCGT
                                 Amp resistance gene
7981   AAAAAGG GAATAAG GGCGACA CGGAAAT GTTGAAT ACTCATA CTCTTCC TTTTTCA ATATTAT TGAAGCA
       TTTTTCC CTTATTC CCGCTGT GCCTTTA CAACTTA TGAGTAT GAGAAGG AAAAAGT TATAATA ACTTCGT
                                                         ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
              Amp resistance gene
8051   TTTATCA GGGTTAT TGTCTCA TGAGCGG ATACATA TTTGAAT GTATTTA GAAAAAT AAACAAA TAGGGGT
       AAATAGT CCCAATA ACAGAGT ACTCGCC TATGTAT AAACTTA CATAAAT CTTTTTA TTTGTTT ATCCCCA
8121   TCCGCGC ACATTTC CCCGAAA AGTGCCA CCTGACG TCTAAGA AACCATT ATTATCA TGACATT AACCTAT
       AGGCGCG TGTAAAG GGGCTTT TCACGGT GGACTGC AGATTCT TTGGTAA TAATAGT ACTGTAA TTGGATA
8191   AAAAATA GGCGTAT CACGAG
       TTTTTAT CCGCATA GTGCTC
```

FIGURE 2A

```
                    1                                                  50
          mCEA(6D)  ATGGAGTCTC CCTCGGCCCC TCCCCACAGA TGGTGCATCC CCTGGCAGAG
mCEA(6D,1st&2nd)    ATGGAGTCTC CCTCGGCCCC TCCCCACAGA TGGTGCATCC CCTGGCAGAG

                    51                                                100
          mCEA(6D)  GCTCCTGCTC ACAGCCTCAC TTCTAACCTT CTGGAACCCG CCCACCACTG
mCEA(6D,1st&2nd)    GCTCCTGCTC ACAGCCTCAC TTCTAACCTT CTGGAACCCG CCCACCACTG

                    101                                               150
          mCEA(6D)  CCAAGCTCAC TATTGAATCC ACGCCGTTCA ATGTCGCAGA GGGGAAGGAG
mCEA(6D,1st&2nd)    CCAAGCTCAC TATTGAATCC ACGCCGTTCA ATGTCGCAGA GGGGAAGGAG

                    151                                               200
          mCEA(6D)  GTGCTTCTAC TTGTCCACAA TCTGCCCCAG CATCTTTTTG GCTACAGCTG
mCEA(6D,1st&2nd)    GTGCTTCTAC TTGTCCACAA TCTGCCCCAG CATCTTTTTG GCTACAGCTG

                    201                                               250
          mCEA(6D)  GTACAAAGGT GAAAGAGTGG ATGGCAACCG TCAAATTATA GGATATGTAA
mCEA(6D,1st&2nd)    GTACAAAGGT GAAAGAGTGG ATGGCAACCG TCAAATTATA GGATATGTAA

                    251                                               300
          mCEA(6D)  TAGGAACTCA ACAAGCTACC CCAGGGCCCG CATACAGTGG TCGAGAGATA
mCEA(6D,1st&2nd)    TAGGAACTCA ACAAGCTACC CCAGGGCCCG CATACAGTGG TCGAGAGATA

                    301                                               350
          mCEA(6D)  ATATACCCCA ATGCATCCCT GCTGATCCAG AACATCATCC AGAATGACAC
mCEA(6D,1st&2nd)    ATATACCCCA ATGCATCCCT GCTGATCCAG AACATCATCC AGAATGACAC

                    351                                               400
          mCEA(6D)  AGGATTCTAC ACCCTACACG TCATAAAGTC AGATCTTGTG AATGAAGAAG
mCEA(6D,1st&2nd)    AGGATTCTAC ACCCTACACG TCATAAAGTC AGATCTTGTG AATGAAGAAG

                    401                                               450
          mCEA(6D)  CAACTGGCCA GTTCCGGGTA TACCCGGAGC TGCCCAAGCC CTCCATCTCC
mCEA(6D,1st&2nd)    CAACTGGCCA GTTCCGGGTA TACCCGGAAC TCCCTAAGCC TTCTATTAGC

                    451                                               500
          mCEA(6D)  AGCAACAACT CCAAACCCGT GGAGGACAAG GATGCTGTGG CCTTCACCTG
mCEA(6D,1st&2nd)    TCCAATAATA GTAAGCCTGT CGAAGACAAA GATGCCGTCG CTTTTACATG

                    501                                               550
          mCEA(6D)  TGAACCTGAG ACTCAGGACG CAACCTACCT GTGGTGGGTA AACAATCAGA
mCEA(6D,1st&2nd)    CGAGCCCGAA ACTCAAGACG CAACATATCT CTGGTGGGTG AACAACCAGT

                    551                                               600
          mCEA(6D)  GCCTCCCGGT CAGTCCCAGG CTGCAGCTGT CCAATGGCAA CAGGACCCTC
mCEA(6D,1st&2nd)    CCCTGCCTGT GTCCCCTAGA CTCCAACTCA GCAACGGAAA TAGAACTCTG

                    601                                               650
          mCEA(6D)  ACTCTATTCA ATGTCACAAG AAATGACACA GCAAGCTACA AATGTGAAAC
mCEA(6D,1st&2nd)    ACCCTGTTTA ACGTGACCAG GAACGACACA GCAAGCTACA AATGCGAAAC
```

FIGURE 2B

```
                      651                                                    700
          mCEA(6D)    CCAGAACCCA GTGAGTGCCA GGCGCAGTGA TTCAGTCATC CTGAATGTCC
mCEA(6D,1st&2nd)      CCAAAATCCA GTCAGCGCCA GGAGGTCTGA TTCAGTGATT CTCAACGTGC

                      701                                                    750
          mCEA(6D)    TCTATGGCCC GGATGCCCCC ACCATTTCCC CTCTAAACAC ATCTTACAGA
mCEA(6D,1st&2nd)      TTTACGGACC CGATGCTCCT ACAATCAGCC CTCTAAACAC AAGCTATAGA

                      751                                                    800
          mCEA(6D)    TCAGGGGAAA ATCTGAACCT CTCCTGCCAC GCAGCCTCTA ACCCACCTGC
mCEA(6D,1st&2nd)      TCAGGGGAAA ATCTGAATCT GAGCTGTCAT GCCGCTAGCA ATCCTCCCGC

                      801                                                    850
          mCEA(6D)    ACAGTACTCT TGGTTTGTCA ATGGGACTTT CCAGCAATCC ACCCAAGAGC
mCEA(6D,1st&2nd)      CCAATACAGC TGGTTTGTCA ATGGCACTTT CCAACAGTCC ACCCAGGAAC

                      851                                                    900
          mCEA(6D)    TCTTTATCCC CAACATCACT GTGAATAATA GTGGATCCTA TACGTGCCAA
mCEA(6D,1st&2nd)      TGTTCATTCC CAATATTACC GTGAACAATA GTGGATCCTA CACGTGCCAA

                      901                                                    950
          mCEA(6D)    GCCCATAACT CAGACACTGG CCTCAATAGG ACCACAGTCA CGACGATCAC
mCEA(6D,1st&2nd)      GCTCACAATA GCGACACCGG ACTCAACCGC ACAACCGTGA CGACGATTAC

                      951                                                   1000
          mCEA(6D)    AGTCTATGAG CCACCCAAAC CCTTCATCAC CAGCAACAAC TCCAACCCCG
mCEA(6D,1st&2nd)      CGTGTATGAG CCACCAAAAC CATTCATAAC TAGTAACAAT TCTAACCCAG

                      1001                                                  1050
          mCEA(6D)    TGGAGGATGA GGATGCTGTA GCCTTAACCT GTGAACCTGA GATTCAGAAC
mCEA(6D,1st&2nd)      TTGAGGATGA GGACGCAGTT GCATTAACTT GTGAGCCAGA GATTCAAAAT

                      1051                                                  1100
          mCEA(6D)    ACAACCTACC TGTGGTGGGT AAATAATCAG AGCCTCCCGG TCAGTCCCAG
mCEA(6D,1st&2nd)      ACCACTTATT TATGGTGGGT CAATAACCAA AGTTTGCCGG TTAGCCCACG

                      1101                                                  1150
          mCEA(6D)    GCTGCAGCTG TCCAATGACA ACAGGACCCT CACTCTACTC AGTGTCACAA
mCEA(6D,1st&2nd)      CTTGCAGTTG TCTAATGATA ACCGCACATT GACACTCCTG TCCGTTACTC

                      1151                                                  1200
          mCEA(6D)    GGAATGATGT AGGACCCTAT GAGTGTGGAA TCCAGAACGA ATTAAGTGTT
mCEA(6D,1st&2nd)      GCAATGATGT AGGACCTTAT GAGTGTGGCA TTCAGAATGA ATTATCCGTT

                      1201                                                  1250
          mCEA(6D)    GACCACAGCG ACCCAGTCAT CCTGAATGTC CTCTATGGCC CAGACGACCC
mCEA(6D,1st&2nd)      GATCACTCCG ACCCTGTTAT CCTTAATGTT TTGTATGGCC CAGACGACCC

                      1251                                                  1300
          mCEA(6D)    CACCATTTCC CCCTCATACA CCTATTACCG TCCAGGGGTG AACCTCAGCC
mCEA(6D,1st&2nd)      AACTATATCT CCATCATACA CCTACTACCG TCCCGGCGTG AACTTGAGCC
```

FIGURE 2C

```
                     1301                                                1350
           mCEA(6D)  TCTCCTGCCA TGCAGCCTCT AACCCACCTG CACAGTATTC TTGGCTGATT
 mCEA(6D,1st&2nd)    TTTCTTGCCA TGCAGCATCC AACCCCCCTG CACAGTACTC CTGGCTGATT

                     1351                                                1400
           mCEA(6D)  GATGGGAACA TCCAGCAACA CACACAAGAG CTCTTTATCT CCAACATCAC
 mCEA(6D,1st&2nd)    GATGGAAACA TTCAGCAGCA TACTCAAGAG TTATTTATAA GCAACATAAC

                     1401                                                1450
           mCEA(6D)  TGAGAAGAAC AGCGGACTCT ATACCTGCCA GGCCAATAAC TCAGCCAGTG
 mCEA(6D,1st&2nd)    TGAGAAGAAC AGCGGACTCT ATACTTGCCA GGCCAATAAC TCAGCCAGTG

                     1451                                                1500
           mCEA(6D)  GCCACAGCAG GACTACAGTC AAGACAATCA CAGTCTCTGC GGAGCTGCCC
 mCEA(6D,1st&2nd)    GTCACAGCAG GACTACAGTT AAAACAATAA CTGTTTCCGC GGAGCTGCCC

                     1501                                                1550
           mCEA(6D)  AAGCCCTCCA TCTCCAGCAA CAACTCCAAA CCCGTGGAGG ACAAGGATGC
 mCEA(6D,1st&2nd)    AAGCCCTCCA TCTCCAGCAA CAACTCCAAA CCCGTGGAGG ACAAGGATGC

                     1551                                                1600
           mCEA(6D)  TGTGGCCTTC ACCTGTGAAC CTGAGGCTCA GAACACAACC TACCTGTGGT
 mCEA(6D,1st&2nd)    TGTGGCCTTC ACCTGTGAAC CTGAGGCTCA GAACACAACC TACCTGTGGT

                     1601                                                1650
           mCEA(6D)  GGGTAAATGG TCAGAGCCTC CCAGTCAGTC CCAGGCTGCA GCTGTCCAAT
 mCEA(6D,1st&2nd)    GGGTAAATGG TCAGAGCCTC CCAGTCAGTC CCAGGCTGCA GCTGTCCAAT

                     1651                                                1700
           mCEA(6D)  GGCAACAGGA CCCTCACTCT ATTCAATGTC ACAAGAAATG ACGCAAGAGC
 mCEA(6D,1st&2nd)    GGCAACAGGA CCCTCACTCT ATTCAATGTC ACAAGAAATG ACGCAAGAGC

                     1701                                                1750
           mCEA(6D)  CTATGTATGT GGAATCCAGA ACTCAGTGAG TGCAAACCGC AGTGACCCAG
 mCEA(6D,1st&2nd)    CTATGTATGT GGAATCCAGA ACTCAGTGAG TGCAAACCGC AGTGACCCAG

                     1751                                                1800
           mCEA(6D)  TCACCCTGGA TGTCCTCTAT GGGCCGGACA CCCCCATCAT TTCCCCCCCA
 mCEA(6D,1st&2nd)    TCACCCTGGA TGTCCTCTAT GGGCCGGACA CCCCCATCAT TTCCCCCCCA

                     1801                                                1850
           mCEA(6D)  GACTCGTCTT ACCTTTCGGG AGCGGACCTC AACCTCTCCT GCCACTCGGC
 mCEA(6D,1st&2nd)    GACTCGTCTT ACCTTTCGGG AGCGGACCTC AACCTCTCCT GCCACTCGGC

                     1851                                                1900
           mCEA(6D)  CTCTAACCCA TCCCCGCAGT ATTCTTGGCG TATCAATGGG ATACCGCAGC
 mCEA(6D,1st&2nd)    CTCTAACCCA TCCCCGCAGT ATTCTTGGCG TATCAATGGG ATACCGCAGC

                     1901                                                1950
           mCEA(6D)  AACACACACA AGTTCTCTTT ATCGCCAAAA TCACGCCAAA TAATAACGGG
 mCEA(6D,1st&2nd)    AACACACACA AGTTCTCTTT ATCGCCAAAA TCACGCCAAA TAATAACGGG
```

FIGURE 2D

```
                    1951                                                 2000
          mCEA(6D)  ACCTATGCCT GTTTTGTCTC TAACTTGGCT ACTGGCCGCA ATAATTCCAT
mCEA(6D,1st&2nd)    ACCTATGCCT GTTTTGTCTC TAACTTGGCT ACTGGCCGCA ATAATTCCAT

                    2001                                                 2050
          mCEA(6D)  AGTCAAGAGC ATCACAGTCT CTGCATCTGG AACTTCTCCT GGTCTCTCAG
mCEA(6D,1st&2nd)    AGTCAAGAGC ATCACAGTCT CTGCATCTGG AACTTCTCCT GGTCTCTCAG

                    2051                                                 2100
          mCEA(6D)  CTGGGGCCAC TGTCGGCATC ATGATTGGAG TGCTGGTTGG GGTTGCTCTG
mCEA(6D,1st&2nd)    CTGGGGCCAC TGTCGGCATC ATGATTGGAG TGCTGGTTGG GGTTGCTCTG

                    2101
          mCEA(6D)  ATATAG
mCEA(6D,1st&2nd)    ATATAG
```

FIGURE 3

A. Amino Acid Sequence Comparison of "Wild-Type KSA" (1) and Modified KSA (2)

```
1 MAPPQVLAFGLLLAAATATFAAAQEECVCENYKLAVNCFVNNNRQCQCTSVGAQNTVIC
2 MAPPQVLAFGLLLAAATATFAAAQEECVCENYKLAVNCFVNNNRQCQCTSVGAQNTVIC

1 SKLAAKCLVMKAEMNGSKLGRRAKPEGALQNNDGLYDPDCDESGLFKAKQCNGTSTCWC
2 SKLAAKCLVMKAEMNGSKLGRRAKPEGALQNNDGLYDPDCDESGLFKAKQCNGTSTCWC

1 VNTAGVRRTDKDTEITCSERVRTYWIIIELKHKAREKPYDSKSLRTALQKEITTRYQLD
2 VNTAGVRRTDKDTEITCSERVRTYWIIIELKHKAREKPYDSKSLRTALQKEITTRYQLD

1 PKFITSILYENNVITIDLVQNSSQKTQNDVDIADVAYYFEKDVKGESLFHSKKMDLTVN
2 PKFITSVLYENNVITIDLVQNSSQKTQNDVDIADVAYYFEKDVKGESLFHSKKMDLTVN

1 GEQLDLDPGQTLIYYVDEKAPEFSMQGLKAGVIAVIVVVVIAVVAGIVVLVISRKKRMA
2 GEQLDLDPGQTLIYYVDEKAPEFSMQGLKAGVIAVIVVVVIAVVAGIVVLVISRKKRMA

1 KYEKAEIKEMGEMHRELNA
2 KYEKAEIKEMGEMHRELNA
```

B. DNA Sequence of Modified KSA

```
atggcgcccccgcaggtcctcgcgttcgggcttctgcttgccgcggcgacggcgactttgccgcagctcaggaa
gaatgtgtctgtgaaaactacaagctggccgtaaactgctttgtgaataataatcgtcaatgccagtgtacttca
gttggtgcacaaaatactgtcatttgctcaaagctggctgccaaatgtttggtgatgaaggcagaaatgaatggc
tcaaaacttgggagaagagcaaaacctgaaggggccctccagaacaatgatgggctttatgatcctgactgcgat
gagagcgggctctttaaggccaagcagtgcaacggcacctccacgtgctggtgtgtgaacactgctggggtcaga
agaacagacaaggacactgaaataacctgctctgagcgagtgagaacctactggatcatcattgaactaaaacac
aaagcaagagaaaaaccttatgatagtaaaagtttgcggactgcacttcagaaggagatcacaacgcgttatcaa
ctggatccaaaatttatcacgagtgtgttgtatgagaataatgttatcactattgatctggttcaaaattcttct
caaaaaactcagaatgatgtggacatagctgatgtggcttattattttgaaaaagatgttaaaggtgaatccttg
tttcattctaagaaaatggacctgacagtaaatggggaacaactggatctggatcctggtcaaactttaatttat
tatgttgatgaaaaagcacctgaattctcaatgcagggtctaaaagctggtgttattgctgttattgtggttgtg
gtgatagcagttgttgctggaattgttgtgctggttatttccagaaagaagagaatggcaaagtatgagaaggct
gagataaaggagatgggtgagatgcatagggaactcaatgcataa
```

FIGURE 4A

Construction of Modified KSA Plasmid

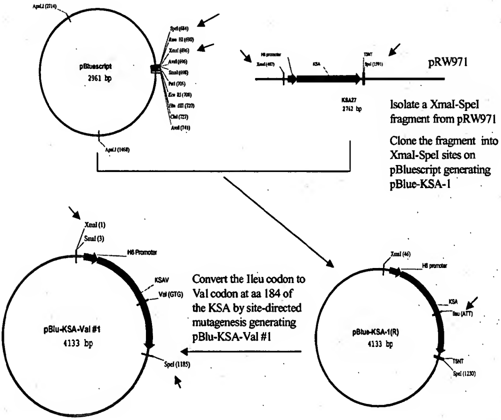


FIGURE 4B
Construction of Modified KSA Plasmid

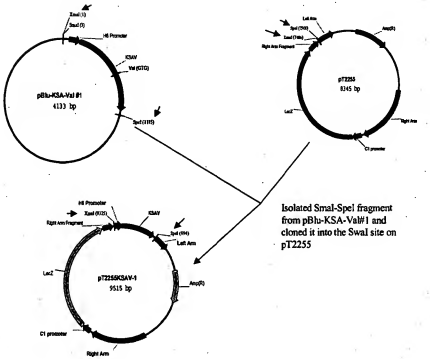
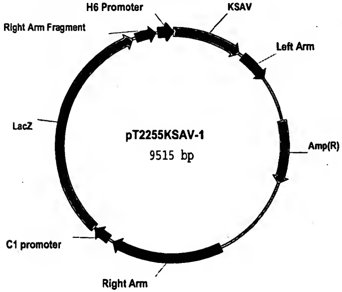


FIGURE 5

A. Plasmid Map of Modified KSA Expression Vector



B. DNA Sequence of Modified KSA Expression Vector

Promoter H6 for KSAV	**9930-9515**
KSAV	**1-945**
Left arm	**1002-1422**
Right arm	**4070-5590**
Right arm fragment	**9012-9299**

```
        MetAlaProPro GlnValLeu AlaPheGly LeuLeuLeuAla AlaAlaThr·
    1   ATGGCGCCCC CGCAGGTCCT CGCGTTCGGG CTTCTGCTTG CCGCGGCGAC
        TACCGCGGGG GCGTCCAGGA GCGCAAGCCC GAAGACGAAC GGCGCCGCTG
        .AlaThrPhe AlaAlaAlaGln GluGluCys ValCysGlu AsnTyrLysLeu·
   51   GGCGACTTTT GCCGCAGCTC AGGAAGAATG TGTCTGTGAA AACTACAAGC
        CCGCTGAAAA CGGCGTCGAG TCCTTCTTAC ACAGACACTT TTGATGTTCG
        ..AlaValAsn CysPheVal AsnAsnAsnArg GlnCysGln CysThrSer
  101   TGGCCGTAAA CTGCTTTGTG AATAATAATC GTCAATGCCA GTGTACTTCA
        ACCGGCATTT GACGAAACAC TTATTATTAG CAGTTACGGT CACATGAAGT
        ValGlyAlaGln AsnThrVal IleCysSer LysLeuAlaAla LysCysLeu·
  151   GTTGGTGCAC AAAATACTGT CATTTGCTCA AAGCTGGCTG CCAAATGTTT
        CAACCACGTG TTTTATGACA GTAAACGAGT TTCGACCGAC GGTTTACAAA
        .ValMetLys AlaGluMetAsn GlySerLys LeuGlyArg ArgAlaLysPro·
  201   GGTGATGAAG GCAGAAATGA ATGGCTCAAA ACTTGGGAGA AGAGCAAAAC
        CCACTACTTC CGTCTTTACT TACCGAGTTT TGAACCCTCT TCTCGTTTTG
        ..GluGlyAla LeuGlnAsn AsnAspGlyLeu TyrAspPro AspCysAsp
  251   CTGAAGGGGC CCTCCAGAAC AATGATGGGC TTTATGATCC TGACTGCGAT
        GACTTCCCCG GGAGGTCTTG TTACTACCCG AAATACTAGG ACTGACGCTA
        GluSerGlyLeu PheLysAla LysGlnCys AsnGlyThrSer ThrCysTrp·
  301   GAGAGCGGGC TCTTTAAGGC CAAGCAGTGC AACGGCACCT CCACGTGCTG
        CTCTCGCCCG AGAAATTCCG GTTCGTCACG TTGCCGTGGA GGTGCACGAC
        .CysValAsn ThrAlaGlyVal ArgArgThr AspLysAsp ThrGluIleThr·
  351   GTGTGTGAAC ACTGCTGGGG TCAGAAGAAC AGACAAGGAC ACTGAAATAA
        CACACACTTG TGACGACCCC AGTCTTCTTG TCTGTTCCTG TGACTTTATT
```

```
      ..CysSerGlu ArgValArg ThrTyrTrpIle IleIleGlu LeuLysHis
 401  CCTGCTCTGA GCGAGTGAGA ACCTACTGGA TCATCATTGA ACTAAAACAC
      GGACGAGACT CGCTCACTCT TGGATGACCT AGTAGTAACT TGATTTTGTG
      LysAlaArgGlu LysProTyr AspSerLys SerLeuArgThr AlaLeuGln·
 451  AAAGCAAGAG AAAAACCTTA TGATAGTAAA AGTTTGCGGA CTGCACTTCA
      TTTCGTTCTC TTTTTGGAAT ACTATCATTT TCAAACGCCT GACGTGAAGT
      .LysGluIle ThrThrArgTyr GlnLeuAsp ProLysPhe IleThrSerVal·
 501  GAAGGAGATC ACAACGCGTT ATCAACTGGA TCCAAAATTT ATCACGAGTG
      CTTCCTCTAG TGTTGCGCAA TAGTTGACCT AGGTTTTAAA TAGTGCTCAC
      ..LeuTyrGlu AsnAsnVal IleThrIleAsp LeuValGln AsnSerSer
 551  TGTTGTATGA GAATAATGTT ATCACTATTG ATCTGGTTCA AAATTCTTCT
      ACAACATACT CTTATTACAA TAGTGATAAC TAGACCAAGT TTTAAGAAGA
      GlnLysThrGln AsnAspVal AspIleAla AspValAlaTyr TyrPheGlu·
 601  CAAAAAACTC AGAATGATGT GGACATAGCT GATGTGGCTT ATTATTTTGA
      GTTTTTTGAG TCTTACTACA CCTGTATCGA CTACACCGAA TAATAAAACT
      .LysAspVal LysGlyGluSer LeuPheHis SerLysLys MetAspLeuThr·
 651  AAAAGATGTT AAAGGTGAAT CCTTGTTTCA TTCTAAGAAA ATGGACCTGA
      TTTTCTACAA TTTCCACTTA GGAACAAAGT AAGATTCTTT TACCTGGACT
      ..ValAsnGly GluGlnLeu AspLeuAspPro GlyGlnThr LeuIleTyr
 701  CAGTAAATGG GGAACAACTG GATCTGGATC CTGGTCAAAC TTTAATTTAT
      GTCATTTACC CCTTGTTGAC CTAGACCTAG GACCAGTTTG AAATTAAATA
      TyrValAspGlu LysAlaPro GluPheSer MetGlnGlyLeu LysAlaGly·
 751  TATGTTGATG AAAAAGCACC TGAATTCTCA ATGCAGGGTC TAAAAGCTGG
      ATACAACTAC TTTTTCGTGG ACTTAAGAGT TACGTCCCAG ATTTTCGACC
      .ValIleAla ValIleValVal ValValIle AlaValVal AlaGlyIleVal·
 801  TGTTATTGCT GTTATTGTGG TTGTGGTGAT AGCAGTTGTT GCTGGAATTG
      ACAATAACGA CAATAACACC AACACCACTA TCGTCAACAA CGACCTTAAC
      ..ValLeuVal IleSerArg LysLysArgMet AlaLysTyr GluLysAla
 851  TTGTGCTGGT TATTTCCAGA AAGAAGAGAA TGGCAAAGTA TGAGAAGGCT
      AACACGACCA ATAAAGGTCT TTCTTCTCTT ACCGTTTCAT ACTCTTCCGA
      GluIleLysGlu MetGlyGlu MetHisArg GluLeuAsnAla ***
 901  GAGATAAAGG AGATGGGTGA GATGCATAGG GAACTCAATG CATAAGAAGC
      CTCTATTTCC TCTACCCACT CTACGTATCC CTTGAGTTAC GTATTCTTCG
 951  TTATCGATAC CGTCGACCTC GAGGAATTCT TTTTATTGAT TAACTAGTTA
      AATAGCTATG GCAGCTGGAG CTCCTTAAGA AAAATAACTA ATTGATCAAT
1001  ATCACGGCCG CTTATAAAGA TCTAAAATGC ATAATTTCTA AATAATGAAA
      TAGTGCCGGC GAATATTTCT AGATTTTACG TATTAAAGAT TTATTACTTT
1051  AAAAGTACA TCATGAGCAA CGCGTTAGTA TATTTTACAA TGGAGATTAA
      TTTTTCATGT AGTACTCGTT GCGCAATCAT ATAAAATGTT ACCTCTAATT
1101  CGCTCTATAC CGTTCTATGT TTATTGATTC AGATGATGTT TTAGAAAAGA
      GCGAGATATG GCAAGATACA AATAACTAAG TCTACTACAA AATCTTTTCT
1151  AAGTTATTGA ATATGAAAAC TTTAATGAAG ATGAAGATGA CGACGATGAT
      TTCAATAACT TATACTTTTG AAATTACTTC TACTTCTACT GCTGCTACTA
1201  TATTGTTGTA AATCTGTTTT AGATGAAGAA GATGACGCGC TAAAGTATAC
      ATAACAACAT TTAGACAAAA TCTACTTCTT CTACTGCGCG ATTTCATATG
1251  TATGGTTACA AAGTATAAGT CTATACTACT AATGGCGACT TGTGCAAGAA
      ATACCAATGT TTCATATTCA GATATGATGA TTACCGCTGA ACACGTTCTT
1301  GGTATAGTAT AGTGAAAATG TTGTTAGATT ATGATTATGA AAAACCAAAT
      CCATATCATA TCACTTTTAC AACAATCTAA TACTAATACT TTTTGGTTTA
1351  AAATCAGATC CATATCTAAA GGTATCTCCT TTGCACATAA TTTCATCTAT
      TTTAGTCTAG GTATAGATTT CCATAGAGGA AACGTGTATT AAAGTAGATA
1401  TCCTAGTTTA GAATACCTGC AGCCAAGCTT GGCACTGGCC GTCGTTTTAC
      AGGATCAAAT CTTATGGACG TCGGTTCGAA CCGTGACCGG CAGCAAAATG
1451  AACGTCGTGA CTGGGAAAAC CCTGGCGTTA CCCAACTTAA TCGCCTTGCA
      TTGCAGCACT GACCCTTTTG GGACCGCAAT GGGTTGAATT AGCGGAACGT
1501  GCACATCCCC CTTTCGCCAG CTGGCGTAAT AGCGAAGAGG CCCGCACCGA
      CGTGTAGGGG GAAAGCGGTC GACCGCATTA TCGCTTCTCC GGGCGTGGCT
1551  TCGCCCTTCC CAACAGTTGC GCAGCCTGAA TGGCGAATGG CGCCTGATGC
```

```
        AGCGGGAAGG GTTGTCAACG CGTCGGACTT ACCGCTTACC GCGGACTACG
1601    GGTATTTTCT CCTTACGCAT CTGTGCGGTA TTTCACACCG CATATGGTGC
        CCATAAAAGA GGAATGCGTA GACACGCCAT AAAGTGTGGC GTATACCACG
1651    ACTCTCAGTA CAATCTGCTC TGATGCCGCA TAGTTAAGCC AGCCCCGACA
        TGAGAGTCAT GTTAGACGAG ACTACGGCGT ATCAATTCGG TCGGGGCTGT
1701    CCCGCCAACA CCCGCTGACG CGCCCTGACG GGCTTGTCTG CTCCCGGCAT
        GGGCGGTTGT GGGCGACTGC GCGGGACTGC CCGAACAGAC GAGGGCCGTA
1751    CCGCTTACAG ACAAGCTGTG ACCGTCTCCG GGAGCTGCAT GTGTCAGAGG
        GGCGAATGTC TGTTCGACAC TGGCAGAGGC CCTCGACGTA CACAGTCTCC
1801    TTTTCACCGT CATCACCGAA ACGCGCGAGA CGAAAGGGCC TCGTGATACG
        AAAAGTGGCA GTAGTGGCTT TGCGCGCTCT GCTTTCCCGG AGCACTATGC
1851    CCTATTTTTA TAGGTTAATG TCATGATAAT AATGGTTTCT TAGACGTCAG
        GGATAAAAAT ATCCAATTAC AGTACTATTA TTACCAAAGA ATCTGCAGTC
1901    GTGGCACTTT TCGGGGAAAT GTGCGCGGAA CCCCTATTTG TTTATTTTTC
        CACCGTGAAA AGCCCCTTTA CACGCGCCTT GGGGATAAAC AAATAAAAAG
1951    TAAATACATT CAAATATGTA TCCGCTCATG AGACAATAAC CCTGATAAAT
        ATTTATGTAA GTTTATACAT AGGCGAGTAC TCTGTTATTG GGACTATTTA
2001    GCTTCAATAA TATTGAAAAA GGAAGAGTAT GAGTATTCAA CATTTCCGTG
        CGAAGTTATT ATAACTTTTT CCTTCTCATA CTCATAAGTT GTAAAGGCAC
2051    TCGCCCTTAT TCCCTTTTTT GCGGCATTTT GCCTTCCTGT TTTTGCTCAC
        AGCGGGAATA AGGGAAAAAA CGCCGTAAAA CGGAAGGACA AAAACGAGTG
2101    CCAGAAACGC TGGTGAAAGT AAAAGATGCT GAAGATCAGT TGGGTGCACG
        GGTCTTTGCG ACCACTTTCA TTTTCTACGA CTTCTAGTCA ACCCACGTGC
2151    AGTGGGTTAC ATCGAACTGG ATCTCAACAG CGGTAAGATC CTTGAGAGTT
        TCACCCAATG TAGCTTGACC TAGAGTTGTC GCCATTCTAG GAACTCTCAA
2201    TTCGCCCCGA AGAACGTTTT CCAATGATGA GCACTTTTAA AGTTCTGCTA
        AAGCGGGGCT TCTTGCAAAA GGTTACTACT CGTGAAAATT TCAAGACGAT
2251    TGTGGCGCGG TATTATCCCG TATTGACGCC GGGCAAGAGC AACTCGGTCG
        ACACCGCGCC ATAATAGGGC ATAACTGCGG CCCGTTCTCG TTGAGCCAGC
2301    CCGCATACAC TATTCTCAGA ATGACTTGGT TGAGTACTCA CCAGTCACAG
        GGCGTATGTG ATAAGAGTCT TACTGAACCA ACTCATGAGT GGTCAGTGTC
2351    AAAAGCATCT TACGGATGGC ATGACAGTAA GAGAATTATG CAGTGCTGCC
        TTTTCGTAGA ATGCCTACCG TACTGTCATT CTCTTAATAC GTCACGACGG
2401    ATAACCATGA GTGATAACAC TGCGGCCAAC TTACTTCTGA CAACGATCGG
        TATTGGTACT CACTATTGTG ACGCCGGTTG AATGAAGACT GTTGCTAGCC
2451    AGGACCGAAG GAGCTAACCG CTTTTTTGCA CAACATGGGG GATCATGTAA
        TCCTGGCTTC CTCGATTGGC GAAAAAACGT GTTGTACCCC CTAGTACATT
2501    CTCGCCTTGA TCGTTGGGAA CCGGAGCTGA ATGAAGCCAT ACCAAACGAC
        GAGCGGAACT AGCAACCCTT GGCCTCGACT TACTTCGGTA TGGTTTGCTG
2551    GAGCGTGACA CCACGATGCC TGTAGCAATG GCAACAACGT TGCGCAAACT
        CTCGCACTGT GGTGCTACGG ACATCGTTAC CGTTGTTGCA ACGCGTTTGA
2601    ATTAACTGGC GAACTACTTA CTCTAGCTTC CCGGCAACAA TTAATAGACT
        TAATTGACCG CTTGATGAAT GAGATCGAAG GGCCGTTGTT AATTATCTGA
2651    GGATGGAGGC GGATAAAGTT GCAGGACCAC TTCTGCGCTC GGCCCTTCCG
        CCTACCTCCG CCTATTTCAA CGTCCTGGTG AAGACGCGAG CCGGGAAGGC
2701    GCTGGCTGGT TTATTGCTGA TAAATCTGGA GCCGGTGAGC GTGGGTCTCG
        CGACCGACCA AATAACGACT ATTTAGACCT CGGCCACTCG CACCCAGAGC
2751    CGGTATCATT GCAGCACTGG GGCCAGATGG TAAGCCCTCC CGTATCGTAG
        GCCATAGTAA CGTCGTGACC CCGGTCTACC ATTCGGGAGG GCATAGCATC
2801    TTATCTACAC GACGGGGAGT CAGGCAACTA TGGATGAACG AAATAGACAG
        AATAGATGTG CTGCCCCTCA GTCCGTTGAT ACCTACTTGC TTTATCTGTC
2851    ATCGCTGAGA TAGGTGCCTC ACTGATTAAG CATTGGTAAC TGTCAGACCA
        TAGCGACTCT ATCCACGGAG TGACTAATTC GTAACCATTG ACAGTCTGGT
2901    AGTTTACTCA TATATACTTT AGATTGATTT AAAACTTCAT TTTTAATTTA
        TCAAATGAGT ATATATGAAA TCTAACTAAA TTTTGAAGTA AAAATTAAAT
2951    AAAGGATCTA GGTGAAGATC CTTTTTGATA ATCTCATGAC CAAAATCCCT
        TTTCCTAGAT CCACTTCTAG GAAAAACTAT TAGAGTACTG GTTTTAGGGA
3001    TAACGTGAGT TTTCGTTCCA CTGAGCGTCA GACCCCGTAG AAAAGATCAA
```

```
      ATTGCACTCA AAAGCAAGGT GACTCGCAGT CTGGGGCATC TTTTCTAGTT
3051  AGGATCTTCT TGAGATCCTT TTTTTCTGCG CGTAATCTGC TGCTTGCAAA
      TCCTAGAAGA ACTCTAGGAA AAAAAGACGC GCATTAGACG ACGAACGTTT
3101  CAAAAAAACC ACCGCTACCA GCGGTGGTTT GTTTGCCGGA TCAAGAGCTA
      GTTTTTTTGG TGGCGATGGT CGCCACCAAA CAAACGGCCT AGTTCTCGAT
3151  CCAACTCTTT TTCCGAAGGT AACTGGCTTC AGCAGAGCGC AGATACCAAA
      GGTTGAGAAA AAGGCTTCCA TTGACCGAAG TCGTCTCGCG TCTATGGTTT
3201  TACTGTCCTT CTAGTGTAGC CGTAGTTAGG CCACCACTTC AAGAACTCTG
      ATGACAGGAA GATCACATCG GCATCAATCC GGTGGTGAAG TTCTTGAGAC
3251  TAGCACCGCC TACATACCTC GCTCTGCTAA TCCTGTTACC AGTGGCTGCT
      ATCGTGGCGG ATGTATGGAG CGAGACGATT AGGACAATGG TCACCGACGA
3301  GCCAGTGGCG ATAAGTCGTG TCTTACCGGG TTGGACTCAA GACGATAGTT
      CGGTCACCGC TATTCAGCAC AGAATGGCCC AACCTGAGTT CTGCTATCAA
3351  ACCGGATAAG GCGCAGCGGT CGGGCTGAAC GGGGGGTTCG TGCACACAGC
      TGGCCTATTC CGCGTCGCCA GCCCGACTTG CCCCCCAAGC ACGTGTGTCG
3401  CCAGCTTGGA GCGAACGACC TACACCGAAC TGAGATACCT ACAGCGTGAG
      GGTCGAACCT CGCTTGCTGG ATGTGGCTTG ACTCTATGGA TGTCGCACTC
3451  CTATGAGAAA GCGCCACGCT TCCCGAAGGG AGAAAGGCGG ACAGGTATCC
      GATACTCTTT CGCGGTGCGA AGGGCTTCCC TCTTTCCGCC TGTCCATAGG
3501  GGTAAGCGGC AGGGTCGGAA CAGGAGAGCG CACGAGGGAG CTTCCAGGGG
      CCATTCGCCG TCCCAGCCTT GTCCTCTCGC GTGCTCCCTC GAAGGTCCCC
3551  GAAACGCCTG GTATCTTTAT AGTCCTGTCG GGTTTCGCCA CCTCTGACTT
      CTTTGCGGAC CATAGAAATA TCAGGACAGC CCAAAGCGGT GGAGACTGAA
3601  GAGCGTCGAT TTTTGTGATG CTCGTCAGGG GGGCGGAGCC TATGGAAAAA
      CTCGCAGCTA AAAACACTAC GAGCAGTCCC CCCGCCTCGG ATACCTTTTT
3651  CGCCAGCAAC GCGGCCTTTT TACGGTTCCT GGCCTTTTGC TGGCCTTTTG
      GCGGTCGTTG CGCCGGAAAA ATGCCAAGGA CCGGAAAACG ACCGGAAAAC
3701  CTCACATGTT CTTTCCTGCG TTATCCCCTG ATTCTGTGGA TAACCGTATT
      GAGTGTACAA GAAAGGACGC AATAGGGGAC TAAGACACCT ATTGGCATAA
3751  ACCGCCTTTG AGTGAGCTGA TACCGCTCGC CGCAGCCGAA CGACCGAGCG
      TGGCGGAAAC TCACTCGACT ATGGCGAGCG GCGTCGGCTT GCTGGCTCGC
3801  CAGCGAGTCA GTGAGCGAGG AAGCGGAAGA GCGCCCAATA CGCAAACCGC
      GTCGCTCAGT CACTCGCTCC TTCGCCTTCT CGCGGGTTAT GCGTTTGGCG
3851  CTCTCCCCGC GCGTTGGCCG ATTCATTAAT GCAGCTGGCA CGACAGGTTT
      GAGAGGGGCG CGCAACCGGC TAAGTAATTA CGTCGACCGT GCTGTCCAAA
3901  CCCGACTGGA AAGCGGGCAG TGAGCGCAAC GCAATTAATG TGAGTTAGCT
      GGGCTGACCT TTCGCCCGTC ACTCGCGTTG CGTTAATTAC ACTCAATCGA
3951  CACTCATTAG GCACCCCAGG CTTTACACTT TATGCTTCCG GCTCGTATGT
      GTGAGTAATC CGTGGGGTCC GAAATGTGAA ATACGAAGGC CGAGCATACA
4001  TGTGTGGAAT TGTGAGCGGA TAACAATTTC ACACAGGAAA CAGCTATGAC
      ACACACCTTA ACACTCGCCT ATTGTTAAAG TGTGTCCTTT GTCGATACTG
4051  CATGATTACG AATTGAATTG CGGCCGCAAT TCTGAATGTT AAATGTTATA
      GTACTAATGC TTAACTTAAC GCCGGCGTTA AGACTTACAA TTTACAATAT
4101  CTTTGGATGA AGCTATAAAT ATGCATTGGA AAAATAATCC ATTTAAAGAA
      GAAACCTACT TCGATATTTA TACGTAACCT TTTTATTAGG TAAATTTCTT
4151  AGGATTCAAA TACTACAAAA CCTAAGCGAT AATATGTTAA CTAAGCTTAT
      TCCTAAGTTT ATGATGTTTT GGATTCGCTA TTATACAATT GATTCGAATA
4201  TCTTAACGAC GCTTTAAATA TACACAAATA AACATAATTT TTGTATAACC
      AGAATTGCTG CGAAATTTAT ATGTGTTTAT TTGTATTAAA AACATATTGG
4251  TAACAAATAA CTAAAACATA AAAATAATAA AAGGAAATGT AATATCGTAA
      ATTGTTTATT GATTTTGTAT TTTTATTATT TTCCTTTACA TTATAGCATT
4301  TTATTTTACT CAGGAATGGG GTTAAATATT TATATCACGT GTATATCTAT
      AATAAAATGA GTCCTTACCC CAATTTATAA ATATAGTGCA CATATAGATA
4351  ACTGTTATCG TATACTCTTT ACAATTACTA TTACGAATAT GCAAGAGATA
      TGACAATAGC ATATGAGAAA TGTTAATGAT AATGCTTATA CGTTCTCTAT
4401  ATAAGATTAC GTATTTAAGA GAATCTTGTC ATGATAATTG GGTACGACAT
      TATTCTAATG CATAAATTCT CTTAGAACAG TACTATTAAC CCATGCTGTA
4451  AGTGATAAAT GCTATTTCGC ATCGTTACAT AAAGTCAGTT GGAAAGATGG
```

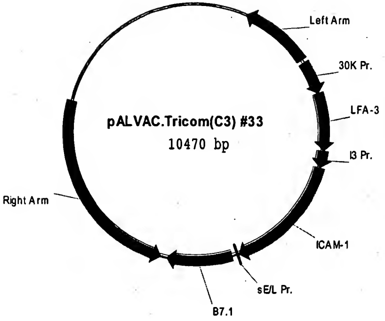
TCACTATTTA CGATAAAGCG TAGCAATGTA TTTCAGTCAA CCTTTCTACC
4501 ATTTGACAGA TGTAACTTAA TAGGTGCAAA AATGTTAAAT AACAGCATTC
TAAACTGTCT ACATTGAATT ATCCACGTTT TTACAATTTA TTGTCGTAAG
4551 TATCGGAAGA TAGGATACCA GTTATATTAT ACAAAAATCA CTGGTTGGAT
ATAGCCTTCT ATCCTATGGT CAATATAATA TGTTTTTAGT GACCAACCTA
4601 AAAACAGATT CTGCAATATT CGTAAAAGAT GAAGATTACT GCGAATTTGT
TTTTGTCTAA GACGTTATAA GCATTTTCTA CTTCTAATGA CGCTTAAACA
4651 AAACTATGAC AATAAAAAGC CATTTATCTC AACGACATCG TGTAATTCTT
TTTGATACTG TTATTTTTCG GTAAATAGAG TTGCTGTAGC ACATTAAGAA
4701 CCATGTTTTA TGTATGTGTT TCAGATATTA TGAGATTACT ATAAACTTTT
GGTACAAAAT ACATACACAA AGTCTATAAT ACTCTAATGA TATTTGAAAA
4751 TGTATACTTA TATTCCGTAA ACTATATTAA TCATGAAGAA AATGAAAAAG
ACATATGAAT ATAAGGCATT TGATATAATT AGTACTTCTT TTACTTTTTC
4801 TATAGAAGCT GTTCACGAGC GGTTGTTGAA AACAACAAAA TTATACATTC
ATATCTTCGA CAAGTGCTCG CCAACAACTT TTGTTGTTTT AATATGTAAG
4851 AAGATGGCTT ACATATACGT CTGTGAGGCT ATCATGGATA ATGACAATGC
TTCTACCGAA TGTATATGCA GACACTCCGA TAGTACCTAT TACTGTTACG
4901 ATCTCTAAAT AGGTTTTTGG ACAATGGATT CGACCCTAAC ACGGAATATG
TAGAGATTTA TCCAAAAACC TGTTACCTAA GCTGGGATTG TGCCTTATAC
4951 GTACTCTACA ATCTCCTCTT GAAATGGCTG TAATGTTCAA GAATACCGAG
CATGAGATGT TAGAGGAGAA CTTTACCGAC ATTACAAGTT CTTATGGCTC
5001 GCTATAAAAA TCTTGATGAG GTATGGAGCT AAACCTGTAG TTACTGAATG
CGATATTTTT AGAACTACTC CATACCTCGA TTTGGACATC AATGACTTAC
5051 CACAACTTCT TGTCTGCATG ATGCGGTGTT GAGAGACGAC TACAAAATAG
GTGTTGAAGA ACAGACGTAC TACGCCACAA CTCTCTGCTG ATGTTTTATC
5101 TGAAAGATCT GTTGAAGAAT AACTATGTAA ACAATGTTCT TTACAGCGGA
ACTTTCTAGA CAACTTCTTA TTGATACATT TGTTACAAGA AATGTCGCCT
5151 GGCTTTACTC CTTTGTGTTT GGCAGCTTAC CTTAACAAAG TTAATTTGGT
CCGAAATGAG GAAACACAAA CCGTCGAATG GAATTGTTTC AATTAAACCA
5201 TAAACTTCTA TTGGCTCATT CGGCGGATGT AGATATTTCA AACACGGATC
ATTTGAAGAT AACCGAGTAA GCCGCCTACA TCTATAAAGT TTGTGCCTAG
5251 GGTTAACTCC TCTACATATA GCCGTATCAA ATAAAAATTT AACAATGGTT
CCAATTGAGG AGATGTATAT CGGCATAGTT TATTTTTAAA TTGTTACCAA
5301 AAACTTCTAT TGAACAAAGG TGCTGATACT GACTTGCTGG ATAACATGGG
TTTGAAGATA ACTTGTTTCC ACGACTATGA CTGAACGACC TATTGTACCC
5351 ATGTACTCCT TTAATGATCG CTGTACAATC TGGAAATATT GAAATATGTA
TACATGAGGA AATTACTAGC GACATGTTAG ACCTTTATAA CTTTATACAT
5401 GCACACTACT TAAAAAAAAT AAAATGTCCA GAACTGGGAA AAATTGATCT
CGTGTGATGA ATTTTTTTTA TTTTACAGGT CTTGACCCTT TTTAACTAGA
5451 TGCCAGCTGT AATTCATGGT AGAAAAGAAG TGCTCAGGCT ACTTTTCAAC
ACGGTCGACA TTAAGTACCA TCTTTTCTTC ACGAGTCCGA TGAAAAGTTG
5501 AAAGGAGCAG ATGTAAACTA CATCTTTGAA AGAAATGGAA AATCATATAC
TTTCCTCGTC TACATTTGAT GTAGAAACTT TCTTTACCTT TTAGTATATG
5551 TGTTTTGGAA TTGATTAAAG AAAGTTACTC TGAGACACAA AAGAGGTAGC
ACAAAACCTT AACTAATTTC TTTCAATGAG ACTCTGTGTT TTCTCCATCG
5601 TGAAGTGGTA CTCTCAAAGG TACGTGACTA ATTAGCTATA AAAAGGATCC
ACTTCACCAT GAGAGTTTCC ATGCACTGAT TAATCGATAT TTTTCCTAGG
5651 TAGAGGATCA TTATTTAACG TAAACTAAAT GGAAAAGCTA TTTACAGGTA
ATCTCCTAGT AATAAATTGC ATTTGATTTA CCTTTTCGAT AAATGTCCAT
5701 CATACGGTGT TTTCTGGAAT CAAATGATTC TGATTTTGAG GATTTTATCA
GTATGCCACA AAAGACCTTA GTTTACTAAG ACTAAAACTC CTAAAATAGT
5751 ATACAATAAT GACAGTGCTA ACTGGTAAAA AAGAAAGCAA ACAATTATCA
TATGTTATTA CTGTCACGAT TGACCATTTT TTCTTTCGTT TGTTAATAGT
5801 TGGCTAACAA TTTTTATTAT ATTTGTAGTA TGCATAGTGG TCTTTACGTT
ACCGATTGTT AAAAATAATA TAAACATCAT ACGTATCACC AGAAATGCAA
5851 TCTTTATTTA AAGTTAATGT GTTAAGATTA AATGGAGTAA TTGGATCCCC
AGAAATAAAT TTCAATTACA CAATTCTAAT TTACCTCATT AACCTAGGGG
5901 CATCGATGGG GAATTCACTG GCCGTCGTTT TACAACGTCG TGACTGGGAA

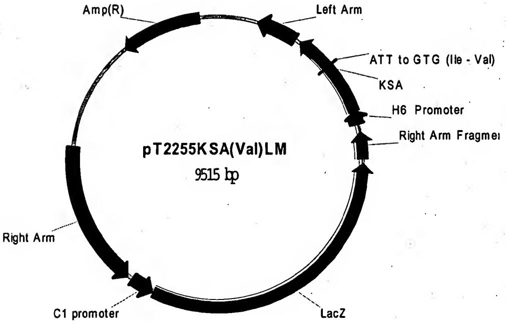
```
      GTAGCTACCC CTTAAGTGAC CGGCAGCAAA ATGTTGCAGC ACTGACCCTT
5951  AACCCTGGCG TTACCCAACT TAATCGCCTT GCAGCACATC CCCCTTTCGC
      TTGGGACCGC AATGGGTTGA ATTAGCGGAA CGTCGTGTAG GGGGAAAGCG
6001  CAGCTGGCGT AATAGCGAAG AGGCCCGCAC CGATCGCCCT TCCCAACAGT
      GTCGACCGCA TTATCGCTTC TCCGGGCGTG GCTAGCGGGA AGGGTTGTCA
6051  TGCGCAGCCT GAATGGCGAA TGGCGCTTTG CCTGGTTTCC GGCACCAGAA
      ACGCGTCGGA CTTACCGCTT ACCGCGAAAC GGACCAAAGG CCGTGGTCTT
6101  GCGGTGCCGG AAAGCTGGCT GGAGTGCGAT CTTCCTGAGG CCGATACTGT
      CGCCACGGCC TTTCGACCGA CCTCACGCTA GAAGGACTCC GGCTATGACA
6151  CGTCGTCCCC TCAAACTGGC AGATGCACGG TTACGATGCG CCCATCTACA
      GCAGCAGGGG AGTTTGACCG TCTACGTGCC AATGCTACGC GGGTAGATGT
6201  CCAACGTAAC CTATCCCATT ACGGTCAATC CGCCGTTTGT TCCCACGGAG
      GGTTGCATTG GATAGGGTAA TGCCAGTTAG GCGGCAAACA AGGGTGCCTC
6251  AATCCGACGG GTTGTTACTC GCTCACATTT AATGTTGATG AAAGCTGGCT
      TTAGGCTGCC CAACAATGAG CGAGTGTAAA TTACAACTAC TTTCGACCGA
6301  ACAGGAAGGC CAGACGCGAA TTATTTTTGA TGGCGTTAAC TCGGCGTTTC
      TGTCCTTCCG GTCTGCGCTT AATAAAAACT ACCGCAATTG AGCCGCAAAG
6351  ATCTGTGGTG CAACGGGCGC TGGGTCGGTT ACGGCCAGGA CAGTCGTTTG
      TAGACACCAC GTTGCCCGCG ACCCAGCCAA TGCCGGTCCT GTCAGCAAAC
6401  CCGTCTGAAT TTGACCTGAG CGCATTTTTA CGCGCCGGAG AAAACCGCCT
      GGCAGACTTA AACTGGACTC GCGTAAAAAT GCGCGGCCTC TTTTGGCGGA
6451  CGCGGTGATG GTGCTGCGTT GGAGTGACGG CAGTTATCTG GAAGATCAGG
      GCGCCACTAC CACGACGCAA CCTCACTGCC GTCAATAGAC CTTCTAGTCC
6501  ATATGTGGCG GATGAGCGGC ATTTTCCGTG ACGTCTCGTT GCTGCATAAA
      TATACACCGC CTACTCGCCG TAAAAGGCAC TGCAGAGCAA CGACGTATTT
6551  CCGACTACAC AAATCAGCGA TTTCCATGTT GCCACTCGCT TTAATGATGA
      GGCTGATGTG TTTAGTCGCT AAAGGTACAA CGGTGAGCGA AATTACTACT
6601  TTTCAGCCGC GCTGTACTGG AGGCTGAAGT TCAGATGTGC GGCGAGTTGC
      AAAGTCGGCG CGACATGACC TCCGACTTCA AGTCTACACG CCGCTCAACG
6651  GTGACTACCT ACGGGTAACA GTTTCTTTAT GGCAGGGTGA AACGCAGGTC
      CACTGATGGA TGCCCATTGT CAAAGAAATA CCGTCCCACT TTGCGTCCAG
6701  GCCAGCGGCA CCGCGCCTTT CGGCGGTGAA ATTATCGATG AGCGTGGTGG
      CGGTCGCCGT GGCGCGGAAA GCCGCCACTT TAATAGCTAC TCGCACCACC
6751  TTATGCCGAT CGCGTCACAC TACGTCTGAA CGTCGAAAAC CCGAAACTGT
      AATACGGCTA GCGCAGTGTG ATGCAGACTT GCAGCTTTTG GGCTTTGACA
6801  GGAGCGCCGA AATCCCGAAT CTCTATCGTG CGGTGGTTGA ACTGCACACC
      CCTCGCGGCT TTAGGGCTTA GAGATAGCAC GCCACCAACT TGACGTGTGG
6851  GCCGACGGCA CGCTGATTGA AGCAGAAGCC TGCGATGTCG GTTTCCGCGA
      CGGCTGCCGT GCGACTAACT TCGTCTTCGG ACGCTACAGC CAAAGGCGCT
6901  GGTGCGGATT GAAAATGGTC TGCTGCTGCT GAACGGCAAG CCGTTGCTGA
      CCACGCCTAA CTTTTACCAG ACGACGACGA CTTGCCGTTC GGCAACGACT
6951  TTCGAGGCGT TAACCGTCAC GAGCATCATC CTCTGCATGG TCAGGTCATG
      AAGCTCCGCA ATTGGCAGTG CTCGTAGTAG GAGACGTACC AGTCCAGTAC
7001  GATGAGCAGA CGATGGTGCA GGATATCCTG CTGATGAAGC AGAACAACTT
      CTACTCGTCT GCTACCACGT CCTATAGGAC GACTACTTCG TCTTGTTGAA
7051  TAACGCCGTG CGCTGTTCGC ATTATCCGAA CCATCCGCTG TGGTACACGC
      ATTGCGGCAC GCGACAAGCG TAATAGGCTT GGTAGGCGAC ACCATGTGCG
7101  TGTGCGACCG CTACGGCCTG TATGTGGTGG ATGAAGCCAA TATTGAAACC
      ACACGCTGGC GATGCCGGAC ATACACCACC TACTTCGGTT ATAACTTTGG
7151  CACGGCATGG TGCCAATGAA TCGTCTGACC GATGATCCGC GCTGGCTACC
      GTGCCGTACC ACGGTTACTT AGCAGACTGG CTACTAGGCG CGACCGATGG
7201  GGCGATGAGC GAACGCGTAA CGCGAATGGT GCAGCGCGAT CGTAATCACC
      CCGCTACTCG CTTGCGCATT GCGCTTACCA CGTCGCGCTA GCATTAGTGG
7251  CGAGTGTGAT CATCTGGTCG CTGGGGAATG AATCAGGCCA CGGCGCTAAT
      GCTCACACTA GTAGACCAGC GACCCCTTAC TTAGTCCGGT GCCGCGATTA
7301  CACGACGCGC TGTATCGCTG GATCAAATCT GTCGATCCTT CCCGCCCGGT
      GTGCTGCGCG ACATAGCGAC CTAGTTTAGA CAGCTAGGAA GGGCGGGCCA
7351  GCAGTATGAA GGCGGCGGAG CCGACACCAC GGCCACCGAT ATTATTTGCC
```

```
      CGTCATACTT CCGCCGCCTC GGCTGTGGTG CCGGTGGCTA TAATAAACGG
7401  CGATGTACGC GCGCGTGGAT GAAGACCAGC CCTTCCCGGC TGTGCCGAAA
      GCTACATGCG CGCGCACCTA CTTCTGGTCG GGAAGGGCCG ACACGGCTTT
7451  TGGTCCATCA AAAAATGGCT TTCGCTACCT GGAGAGACGC GCCCGCTGAT
      ACCAGGTAGT TTTTTACCGA AAGCGATGGA CCTCTCTGCG CGGGCGACTA
7501  CCTTTGCGAA TACGCCCACG CGATGGGTAA CAGTCTTGGC GGTTTCGCTA
      GGAAACGCTT ATGCGGGTGC GCTACCCATT GTCAGAACCG CCAAAGCGAT
7551  AATACTGGCA GGCGTTTCGT CAGTATCCCC GTTTACAGGG CGGCTTCGTC
      TTATGACCGT CCGCAAAGCA GTCATAGGGG CAAATGTCCC GCCGAAGCAG
7601  TGGGACTGGG TGGATCAGTC GCTGATTAAA TATGATGAAA ACGGCAACCC
      ACCCTGACCC ACCTAGTCAG CGACTAATTT ATACTACTTT TGCCGTTGGG
7651  GTGGTCGGCT TACGGCGGTG ATTTTGGCGA TACGCCGAAC GATCGCCAGT
      CACCAGCCGA ATGCCGCCAC TAAAACCGCT ATGCGGCTTG CTAGCGGTCA
7701  TCTGTATGAA CGGTCTGGTC TTTGCCGACC GCACGCCGCA TCCAGCGCTG
      AGACATACTT GCCAGACCAG AAACGGCTGG CGTGCGGCGT AGGTCGCGAC
7751  ACGGAAGCAA AACACCAGCA GCAGTTTTTC CAGTTCCGTT TATCCGGGCA
      TGCCTTCGTT TTGTGGTCGT CGTCAAAAAG GTCAAGGCAA ATAGGCCCGT
7801  AACCATCGAA GTGACCAGCG AATACCTGTT CCGTCATAGC GATAACGAGC
      TTGGTAGCTT CACTGGTCGC TTATGGACAA GGCAGTATCG CTATTGCTCG
7851  TCCTGCACTG GATGGTGGCG CTGGATGGTA AGCCGCTGGC AAGCGGTGAA
      AGGACGTGAC CTACCACCGC GACCTACCAT TCGGCGACCG TTCGCCACTT
7901  GTGCCTCTGG ATGTCGCTCC ACAAGGTAAA CAGTTGATTG AACTGCCTGA
      CACGGAGACC TACAGCGAGG TGTTCCATTT GTCAACTAAC TTGACGGACT
7951  ACTACCGCAG CCGGAGAGCG CCGGGCAACT CTGGCTCACA GTACGCGTAG
      TGATGGCGTC GGCCTCTCGC GGCCCGTTGA GACCGAGTGT CATGCGCATC
8001  TGCAACCGAA CGCGACCGCA TGGTCAGAAG CCGGGCACAT CAGCGCCTGG
      ACGTTGGCTT GCGCTGGCGT ACCAGTCTTC GGCCCGTGTA GTCGCGGACC
8051  CAGCAGTGGC GTCTGGCGGA AAACCTCAGT GTGACGCTCC CCGCCGCGTC
      GTCGTCACCG CAGACCGCCT TTTGGAGTCA CACTGCGAGG GGCGGCGCAG
8101  CCACGCCATC CCGCATCTGA CCACCAGCGA AATGGATTTT TGCATCGAGC
      GGTGCGGTAG GGCGTAGACT GGTGGTCGCT TTACCTAAAA ACGTAGCTCG
8151  TGGGTAATAA GCGTTGGCAA TTTAACCGCC AGTCAGGCTT TCTTTCACAG
      ACCCATTATT CGCAACCGTT AAATTGGCGG TCAGTCCGAA AGAAAGTGTC
8201  ATGTGGATTG GCGATAAAAA ACAACTGCTG ACGCCGCTGC GCGATCAGTT
      TACACCTAAC CGCTATTTTT TGTTGACGAC TGCGGCGACG CGCTAGTCAA
8251  CACCCGTGCA CCGCTGGATA ACGACATTGG CGTAAGTGAA GCGACCCGCA
      GTGGGCACGT GGCGACCTAT TGCTGTAACC GCATTCACTT CGCTGGGCGT
8301  TTGACCCTAA CGCCTGGGTC GAACGCTGGA AGGCGGCGGG CCATTACCAG
      AACTGGGATT GCGGACCCAG CTTGCGACCT TCCGCCGCCC GGTAATGGTC
8351  GCCGAAGCAG CGTTGTTGCA GTGCACGGCA GATACACTTG CTGATGCGGT
      CGGCTTCGTC GCAACAACGT CACGTGCCGT CTATGTGAAC GACTACGCCA
8401  GCTGATTACG ACCGCTCACG CGTGGCAGCA TCAGGGGAAA ACCTTATTTA
      CGACTAATGC TGGCGAGTGC GCACCGTCGT AGTCCCCTTT TGGAATAAAT
8451  TCAGCCGGAA AACCTACCGG ATTGATGGTA GTGGTCAAAT GGCGATTACC
      AGTCGGCCTT TTGGATGGCC TAACTACCAT CACCAGTTTA CCGCTAATGG
8501  GTTGATGTTG AAGTGGCGAG CGATACACCG CATCCGGCGC GGATTGGCCT
      CAACTACAAC TTCACCGCTC GCTATGTGGC GTAGGCCGCG CCTAACCGGA
8551  GAACTGCCAG CTGGCGCAGG TAGCAGAGCG GGTAAACTGG CTCGGATTAG
      CTTGACGGTC GACCGCGTCC ATCGTCTCGC CCATTTGACC GAGCCTAATC
8601  GGCCGCAAGA AAACTATCCC GACCGCCTTA CTGCCGCCTG TTTTGACCGC
      CCGGCGTTCT TTTGATAGGG CTGGCGGAAT GACGGCGGAC AAAACTGGCG
8651  TGGGATCTGC CATTGTCAGA CATGTATACC CCGTACGTCT TCCCGAGCGA
      ACCCTAGACG GTAACAGTCT GTACATATGG GGCATGCAGA AGGGCTCGCT
8701  AAACGGTCTG CGCTGCGGGA CGCGCGAATT GAATTATGGC CCACACCAGT
      TTTGCCAGAC GCGACGCCCT GCGCGCTTAA CTTAATACCG GGTGTGGTCA
8751  GGCGCGGCGA CTTCCAGTTC AACATCAGCC GCTACAGTCA ACAGCAACTG
      CCGCGCCGCT GAAGGTCAAG TTGTAGTCGG CGATGTCAGT TGTCGTTGAC
8801  ATGGAAACCA GCCATCGCCA TCTGCTGCAC GCGGAAGAAG GCACATGGCT
```

```
        TACCTTTGGT CGGTAGCGGT AGACGACGTG CGCCTTCTTC CGTGTACCGA
8851    GAATATCGAC GGTTTCCATA TGGGGATTGG TGGCGACGAC TCCTGGAGCC
        CTTATAGCTG CCAAAGGTAT ACCCCTAACC ACCGCTGCTG AGGACCTCGG
8901    CGTCAGTATC GGCGGAATTC CAGCTGAGCG CCGGTCGCTA CCATTACCAG
        GCAGTCATAG CCGCCTTAAG GTCGACTCGC GGCCAGCGAT GGTAATGGTC
8951    TTGGTCTGGT GTCAAAAATA ATAATAACCG GGCAGGGGGG ATCCGGAGCT
        AACCAGACCA CAGTTTTTAT TATTATTGGC CCGTCCCCCC TAGGCCTCGA
9001    TATCGCAGAT CAATGATCGC TGTACAATCT GGAAATATTG AAATATGTAG
        ATAGCGTCTA GTTACTAGCG ACATGTTAGA CCTTTATAAC TTTATACATC
9051    CACACTACTT AAAAAAAATA AAATGTCCAG AACTGGGAAA AATTGATCTT
        GTGTGATGAA TTTTTTTTAT TTTACAGGTC TTGACCCTTT TTAACTAGAA
9101    GCCAGCTGTA ATTCATGGTA GAAAAGAAGT GCTCAGGCTA CTTTTCAACA
        CGGTCGACAT TAAGTACCAT CTTTTCTTCA CGAGTCCGAT GAAAAGTTGT
9151    AAGGAGCAGA TGTAAACTAC ATCTTTGAAA GAAATGGAAA ATCATATACT
        TTCCTCGTCT ACATTTGATG TAGAAACTTT CTTTACCTTT TAGTATATGA
9201    GTTTTGGAAT TGATTAAAGA AAGTTACTCT GAGACACAAA AGAGGTAGCT
        CAAAACCTTA ACTAATTTCT TTCAATGAGA CTCTGTGTTT TCTCCATCGA
9251    GAAGTGGTAC TCTCAAAGGT ACGTGACTAA TTAGCTATAA AAAGGATCCG
        CTTCACCATG AGAGTTTCCA TGCACTGATT AATCGATATT TTTCCTAGGC
9301    GTACCCTCGA GTCTAGAATC GATCCCGGGT TAATTAATTA GTTATTAGAC
        CATGGGAGCT CAGATCTTAG CTAGGGCCCA ATTAATTAAT CAATAATCTG
9351    AAGGTGAAAA CGAAACTATT TGTAGCTTAA TTAATTAGAG CTTCTTTATT
        TTCCACTTTT GCTTTGATAA ACATCGAATT AATTAATCTC GAAGAAATAA
9401    CTATACTTAA AAAGTGAAAA TAAATACAAA GGTTCTTGAG GGTTGTGTTA
        GATATGAATT TTTCACTTTT ATTTATGTTT CCAAGAACTC CCAACACAAT
9451    AATTGAAAGC GAGAAATAAT CATAAATTAT TTCATTATCG CGATATCCGT
        TTAACTTTCG CTCTTTATTA GTATTTAATA AAGTAATAGC GCTATAGGCA
9501    TAAGTTTGTA TCGTA
        ATTCAAACAT AGCAT
```

FIGURE 6


pALVAC.Tricom(C3) #33
10470 bp
Left Arm
30K Pr.
LFA-3
I3 Pr.
ICAM-1
sE/L Pr.
B7.1
Right Arm


pT2255KSA(Val)LM
9515 bp
Amp(R)
Left Arm
ATT to GTG (Ile - Val)
KSA
H6 Promoter
Right Arm Fragme
LacZ
C1 promoter
Right Arm